The Dissertation Committee for Mingshuang Li Certifies that this is the approved version of the following dissertation:

Perceptual effects of formant enhancement with the factors of phonetic type, listening conditions and language experience of listeners

Committee:

_____
Chang Liu, Supervisor

_____
Craig A. Champlin

_____
Rajka Smiljanic

_____
Julia Campbell

Perceptual effects of formant enhancement with the factors of phonetic type, listening conditions and language experience of listeners

by

Mingshuang Li

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2021

Acknowledgements

Being a PhD is my dream since my childhood, but getting a PhD is never easy. Fortunately, I have received a great deal of help throughout my doctoral program. I really want to thank the people who have ever helped, supported, encouraged, and cared about me. I would have not reached my dream without you.

First, I would like to thank my supervisor, Dr. Chang Liu. for his encouragement, support, and guidance for my academic pursue. Dr. Liu is knowledgeable, inspirational, and patient to students including me, and I hope to be an excellent faculty like him in the future. I would also like to thank Drs. Craig Champlin, Julia Campbell, and Rajka Smiljanic as my Dissertation Committee members. Thank you for your insightful suggestions and comments for my doctoral training and dissertation project.

I feel very fortunate to gain the clinical training in Audiology in my PhD program. I would also like to acknowledge my clinical supervisors, including Dr. Amanda Zappler, Dr. Angela Carey, Joan Balash and Dr. Sangeeta Kamdar at UT, as well as Dr. Yuyan Esther Zhang, Dr. Jingjing Guan, Dr. Brook Johnson, and Dr. Becca Dixon in off-campus clinics.

I also want to give my special thanks to my parents, Mr. Qinkai Li and Mrs. Fengqiu Xu. I am also grateful to my friends I met in Austin, Zhaohe Dai, Ying Hao, Can Xu, Yao Chen, Cissy Chen, Fanyin Chen and Won Soo. In addition, I am grateful to Lujia Yang at Shanghai Jiao Tong University for her help in data collection.

It’s a blessing to have you.

Abstract

Perceptual effects of formant enhancement with the factors of phonetic type, listening conditions and language experience of listeners

Mingshuang Li, Ph.D.

The University of Texas at Austin, 2021

Supervisor: Chang Liu

The second formant (F2) enhancement is a technique that aims to improve speech perception in adverse noise by amplifying the F2 of speech signals. The current study aimed to investigate whether F2 enhancement would improve speech identification with the factors of phonetic type (e.g., vowel and consonant), listening conditions (e.g., speech and nonspeech noise at moderately and very challenging SNRs), and language experience of listeners (e.g., native and nonnative listeners), if any, whether the amount of perceptual benefit was dependent on these factors. Two groups of participants, English native and nonnative listeners, were recruited in this study. Identification of English vowels and consonants with and without F2 enhancement were measured in quiet, long-term speech

shaped noise (LTSSN) and six-talker babble (6-TB) at the signal-to-noise ratios (SNRs) of -10 dB and -15 dB. Overall, significant improvements from F2 enhancement were found in both vowel and consonant identification for both native and nonnative listeners in various listening conditions. Furthermore, greater improvement was found at the SNR of -15 dB than at the SNR of -10 dB, as well as for nonnative listeners than native listeners in vowel identification. Meanwhile, the amount of benefit was generally comparable speech and nonspeech noise. These results indicate that F2 enhancement could improve phonetic identification in noise for native and nonnative listeners, showing a potential as a speech enhancement algorithm in challenging noise.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA ..... Analysis of Variance

CEFS ..... Contrast Enhanced Frequency Shaping

CN ..... Chinese-native

CVC ..... Consonant -Vowel- Consonant

dB ..... Decibels

dB SPL ..... Decibels Sound Pressure Level

EN ..... English-native

F1 ..... First Formant Frequency

F2 ..... Second Formant Frequency

F3 ..... Third Formant Frequency

FFT ..... Fast Fourier Transform

Hz..... Hertz

LTSSN ..... Long-term Speech Shaped Noise

LPC ..... Linear Predictive Coding

MTB ..... Multi-talker Babble

RM ..... Mobile Sound Processor

RMS ..... Root-mean-square

RP ..... Real-time Processor

SJTU .....Shanghai Jiao Tong University

SNR ..... Signal-to-Noise Ratio

UT .....University of Texas at Austin

6-TB ..... Six-talker Babble

Chapter 1: Introduction

Speech perception in adverse noise is challenging for listeners, including those with normal hearing. One of the primary reasons is the masking of background noise on the spectral prominence of target speech, resulting in spectral smearing and/or misallocation. In addition, with the degraded spectral cues (e.g., reduced formant peaks), speech perception becomes more difficult in noise conditions.

Phonetic perception in noise depends on several factors such as phonetic type, listening conditions, and language experience of listeners. The different weights of spectral cues with noise-masking may differ in vowel and consonant identification in noise for phonetic recognition in noise (Parikh & Loizou, 2005). The first and second formants (e.g., F1 and F2) in the steady-state and formant transitions are the primary cues in vowel identification (Hillenbrand, Clark, & Nearey, 2001; Peterson & Barney, 1951; Strange, 1989); thus, degraded formants interfered by background noise significantly affect vowel identification. On the other hand, for consonant identification, formant transitions make important contributions along with other acoustic cues such as the spectra of release burst for stop consonants (Dorman, Studdert-Kennedy, & Raphael, 1977); meanwhile, the weight of formant information in consonant processing may not be as high as in vowel identification. Therefore, listeners can take advantage of preserved cues (e.g., release burst) to identify consonants when the formant cues were degraded by background noise (Dorman et al., 1977; Story & Bunton, 2010).

Listening conditions, including signal-to-noise ratio (SNR) and noise type, are critical to affect phonetic perception. In general, speech recognition increases with SNRs

for a given noise (Liu & Kewley-Port, 2004; Phatak & Allen, 2007; Plomp & Mimpen, 1979b). In addition, the type of noise, e.g., nonspeech and speech noise, is another factor to determine the amount of masking, depending on the masking mechanisms such as energetic and/or informational masking.

Considering listeners' factors, language background is well-known to significantly affect phonetic perception in quiet and noise. Nonnative listeners usually suffer more difficulties in speech perception in adverse noise, which could be attributed to their disadvantages in formant processing of target speech (Flege, Bohn, & Jang, 1997; Kondaurova & Francis, 2008; Liu, Tao, Wang, & Dong, 2012; Mi et al., 2016; Morrison, 2009; Wang, 2006) as well as their reduced capacities against energetic and/or informational masking from background noise (Cooke, Garcia Lecumberri, & Barker, 2008b; Guan et al., 2015; Mi et al., 2013).

Spectral enhancement is considered as a technical solution to improve speech recognition in noise by strengthening spectral cues. Instead of increasing the overall intensity of target speech, this method enhances the amplitudes and peak-to-valley contrasts of spectral peaks (e.g., formants), strengthening spectral cues degraded in background noise. The traditional strategies of spectral enhancement are to expand or/and sharpen the spectral contrasts of formant peaks; however, the perceptual benefits of these algorithms are limited (Bunnell, 1990; Franck, van Kreveld-Bos, Dreschler, & Verschuure, 1999a; Rout, 2006; Stone & Moore, 1992; Summerfield, Foster, Tyler, & Bailey, 1985). Woodall and Liu (2013) proposed a new strategy named F2 enhancement,

focusing on the F2 amplification of target speech without disrupting other frequency areas. Results of several studies suggested that the F2 enhancement could improve formant sensitivity and word recognition in noise for the listeners with normal hearing and hearing loss (Guan & Liu, 2019a; Guan & Liu, 2019b). As the previous studies of F2 enhancement showed a significiant improvement in word recogntion, it was still unknown that at the phonetic level, vowel, consonant, or both received the perceptual benefits at the phonetic level. Thus, the current study examined the effect of F2 enhancement on English phonetic identification in quiet and noise for native and nonnative listeners. In addition, the improvements were compared among these conditions to evaluate whether the effectiveness of F2 enhancement (e.g., the amount of perceptual benefit) depended on these factors.

Chapter 2: Literature Review

There are two areas in this literature review relevant to this study: 1. Speech perception in noise with the factors of phonetic type, noise condition, and listeners' language background; 2. Speech enhancement to improve speech perception in noise.

2.1 Speech perception in noise

Speech communication usually takes place in background noise. In general, background noise has masking effects on acoustic cues of target speech (e.g., formants), leading to difficulties for listeners in speech perception. Previous studies suggested that speech perception in noise depends on several factors such as speech materials, listening conditions, and language experience of listeners. In the present study, vowel and consonant identification was measured in quiet and noise for native and non-native listeners. Thus, the effects of the related factors on speech perception in noise are described below, respectively.

2.1.1 Phonetic perception: vowel and consonant identification

In general, speech sounds are composed of vowels and consonants. Listeners usually suffer difficulties in both vowel and consonant identification in adverse noise, while some studies found more problems in vowel identification than in consonant identification (Mi et al., 2013; Parikh & Loizou, 2005; Tao et al., 2018). Parikh and Loizou (2005) showed that vowel identification was greatly affected by background noise

at the SNR of -5 dB, while consonant recognition remained at a high score at the same SNR. Mi et al. (2013) and Tao et al. (2018) investigated vowel and consonant identification with the similar background noise from moderate (e.g., 0 dB SNR) to very challenging (e.g., -15 dB) SNRs. Combined with the results from the two studies, vowel identification generally suffered more from noise interference than consonant identification at a given SNR. In addition, the consonant-vowel difference in noise increased with the decrease of SNR, e.g., approximately 10% at the SNR of 0 dB and 50% at the SNR of -15 dB. It might be partly attributed to the different perceptual weights of formants, especially F2, on vowel and consonant identification in noise.

Formant is defined as the broad spectral maximum that results from an acoustic resonance of the human vocal tract (Titze et al., 2015; Titze & Martin, 1998), which plays an essential role in both vowel and consonant identification. The frequencies of F1 and F2 are considered as the primary cues to recognize vowels (Hillenbrand et al., 2001; Peterson & Barney, 1951; Strange, 1989). Generally, F1 and F2 frequencies are strong cues to disambiguate and perceive vowel sounds in quiet and adverse listening conditions (Wang, 2017). In addition, formant transitions in the CVC context significantly contributed to vowel identification (Assmann, 1995; Lindblom & Studdert-Kennedy, 1967). Compared with F1 peaks, F2 peaks have lower intensity and are more susceptible to noise interference. Therefore, listeners usually must recognize vowels with fully available F1 and partial F2 in noise (Parikh & Loizou, 2005).

On the other hand, formant information also contributes to consonant recognition through transitions between vowels and consonants in syllables. Formant transition, especially F2 transition, is considered as one of acoustic cues in distinguishing voiced or voiceless consonants, e.g., /p-b/, /t-d/ & /t-g/ (Liberman, Delattre, Cooper, & Gerstman, 1954; Mermelstein, 1978; Stevens & Klatt, 1974), as well as in specifying articulation place, e.g., /b-d-g/ (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Harris, Hoffman, Liberman, Delattre, & Cooper, 1958; Stevens & Blumstein, 1978; Story & Bunton, 2010). However, the weight of formant cues on consonant identification might be generally lower than on vowel identification, e.g., F1 and F2 are the primary cues for vowel identification (Hillenbrand et al., 2001; Peterson & Barney, 1951; Strange, 1989), whereas the formant transitions only might not be sufficient to distinguish consonants (Kewley-Port, 1982). The formant transition contributes to consonant recognition and the information of release burst in voiced-voiceless distinction and place of articulation (Dorman et al., 1977; Stevens & Klatt, 1974; Story & Bunton, 2010). Moreover, the formant transition and release burst showed a trading and reciprocal relationship (Dorman et al., 1977; Story & Bunton, 2010), e.g., where the perceptual weight of one increased, the weight of the other declined. Thus, listeners still could take advantage of the preserved cue of release burst to recognize stop consonants when the formant information was disrupted by adverse background noise.

2.1.2 Noise listening conditions: SNR and noise type

2.1.2.1 SNR

SNR is a primary factor for speech perception in noise. SNR is defined as the target signal power compared with the background noise power, measured in decibel (dB) (Vento & Durrant, 2009). Instead of the overall signal level, speech perception in noise primarily depends on the SNR (Liu & Kewley-Port, 2004; Phatak & Allen, 2007; Plomp & Mimpen, 1979b). Listeners with normal hearing rarely have difficulties in recognizing in noise at the SNRs of 0 dB and above, as formant cues (e.g., F1and F2) are well preserved (Assmann & Summerfield, 2004; Holder, Levin, & Gifford, 2018; Mi et al., 2013). Holder et al. (2018) investigated the effect of SNR on speech perception with the AzBio test, a sentence recognition test in the conditions of quiet and multi-talker babble (MTB) with the SNRs from +10 dB to -10 dB (e.g., 10, 5, 0, -5 and -10 dB). For the young- and middle-aged listeners, the sentence recognition percentage was near 100% at the SNR of 5 dB and above and remained at a high score (90.8%) at the SNR of 0 dB. However, the recognition scores sharply dropped as the SNRs further decreased, e.g., 51.6% at the SNR of -5 dB and 9.0% of -10 dB. The effect of SNR is also presented at phonetic perception. In the study by Mi et al., the percentages of vowel recognition in MTB were about 80% at the SNR of 0 dB, while the score dropped to approximately 55% at the SNR of -9 dB, and further declined to 30% at -15 dB. In the study by Tao et at. (2018), the consonant recognition approximately dropped from 95% at the SNR of 0 dB to 80% at -15 dB. In addition, the effect of SNR also varies across the vowel and consonant categories (Liu & Jin, 2019; Turner, Fabry, Barrett, & Horwitz, 1992). Liu and

Jin (2019) investigated the SNR effect across various vowel categories (e.g., front, middle and back vowels) with psychometric functions. As a result, back and central vowels' identification generally had greater slopes than front vowels, suggesting that front vowels are affected less by the SNR effect than back and central vowels. In addition, Turner et al. (1992) found the detection thresholds of stop consonants were generally lower than vowel (e.g., /a/), while voiceless stops /p, t, k/ showed generally lower thresholds than voiced /p, t, k/, which suggested that the voiceless stops would be detected at a lower SNR compared with voiced stops. In summary, SNR is a primary factor of listening conditions for speech perception in noise, while its perceptual effect also depends on speech materials (e.g., sentence, vowel, and consonant) and phonetic categories. In general, speech recognition improves at 6-9% rate per dB of the SNR increase, depending on speech materials and noise type.

2.1.2.2 Noise type

Noise type is another important factor of listening conditions for speech perception. There are various types of noise in our daily life, e.g., nonspeech and speech noise. There are different mechanisms of masking from nonspeech and speech noise on speech perception, e.g., nonspeech noise primarily contains energetic masking, while speech noise has both energetic and informational masking. Energetic masking refers to the interference when noise and target speech overlap in time and frequency at the peripheral level such that the acoustic information of target speech was reduced (Brungart

& Simpson, 2002; Durlach et al., 2003). Energetic masking usually results in difficulties of phonetic perception by disrupting the spectral cues (e.g., formants) of target speech.

Conversely, formants peaks are masked by background noise when the intensity of noise is higher than formants. In addition, the background noise with lower amplitudes than formant peaks may also degrade the peak-to-valley ratios of formants by filling the valleys in the spectral envelope, namely spectral smearing, resulting in reduced speech intelligibility (Assmann & Summerfield, 2004; Plomp & Mimpen, 1979a; Ter Keurs, Festen, & Plomp, 1992, 1993). Informational masking, primarily originating from the central auditory system, contains any form of masking or interference that cannot be construed as energetic masking (Arbogast, Mason, & Kidd Jr, 2002; Durlach et al., 2003; Kidd, Mason, Richards, Gallun, & Durlach, 2008). Informational masking in speech noise usually degrades the target speech with noise-to-target misallocation (Cooke, Garcia Lecumberri, & Barker, 2008a). Speech noise contains spectrotemporal cues (e.g., formants) like target speech. Therefore, listeners may misallocate the formant information from speech noise into target speech. In addition, informational masking may affect speech recognition with semantic interference, competing attention, and cognitive load (Kahneman, 1973).

Long-term speech-shaped noise (LTSSN) and multi-talker babble (MTB) are usually used in previous studies of speech perception as nonspeech and speech noise (Cooke et al., 2008b; Jin & Liu, 2012; Mi et al., 2013; Tao et al., 2018). The MTB is a type of dynamic speech noise containing speech sounds recorded from one or multiple

talkers, while the LTSSN is stationary nonspeech noise with the average spectrum of MTB. Both LTSSN and MTB include energetic masking on speech perception, while the amount of energetic masking is usually different between them at the same SNRs. However, unlike LTSSN (shown in Figure 1), MTB has more temporal glimpses (shown in Figure 2) that usually reduce the amount of energetic masking. Previous studies found that listeners could take advantage of temporal dips in noise to better speech recognition (Festen & Plomp, 1990; Miller, 1947; Miller & Licklider, 1950). This phenomenon was named as masking release from temporal modulation of noise. In addition, Gustafsson and Arlinger (1994) found that masking release depended on the acoustic features of temporal glimpses, including modulation depth and rate. In general, speech noise contained profound temporal glimpses at middle modulation rates, leading to significant releases of energetic masking. Thus, the energetic masking for MTB with temporal dips is generally lower compared with stationary LTSSN (Cooke et al., 2008).

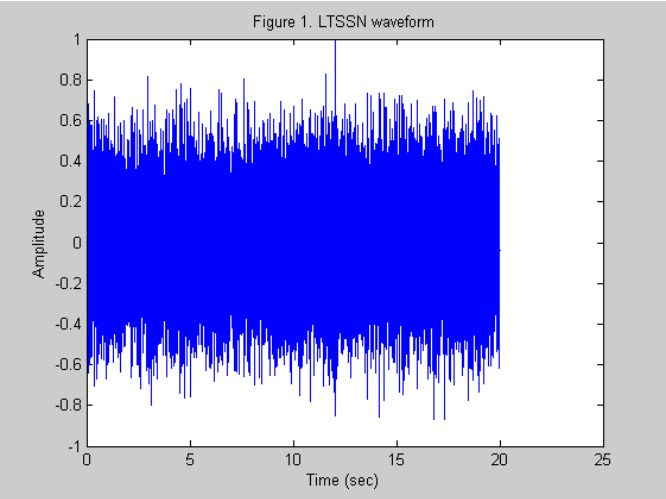

Figure 1. The waveform of long-term speech shaped noise (LTSSN)

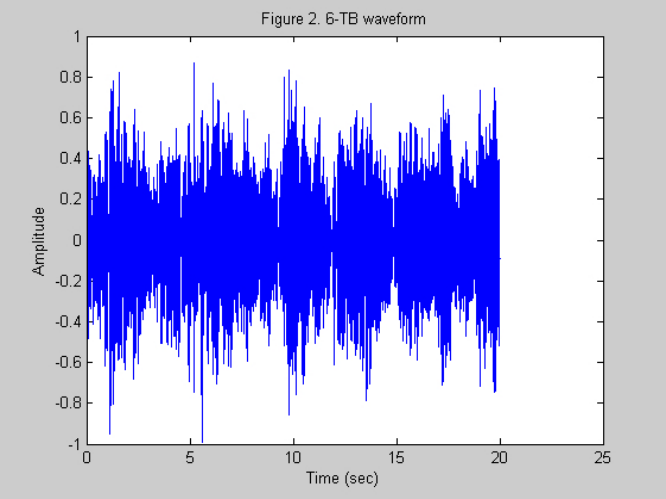

Figure 2. The waveforms of six-talker babble (6-TB)

Compared with LTSSN, listeners usually suffer more informational masking from MTB. Besides the factor of SNR, the amount of informational masking in MTB also depends on the number of talkers and the materials of target speech (e.g., sentence or phoneme). Carhart et al. (1975) investigated the effect of the number of talkers (e.g., 1, 2, 3, 16, 32, 64 and 128 talkers) in MTB on sentence recognition. Results suggested that the highest informational masking occurred in two- and three-talker babbles with the maximal distinguishable semantic information. The masking effect decreased with the increase of the number of talkers in speech babble and stabilized at 64-talker babble. It was suggested that the sentence recognition was mainly affected by semantic interference in informational masking. Meanwhile, in phonetic recognition, the highest informational masking was found in babbles with 6-16 talkers, with the maximum of audible phonetic cues (Simpson & Cooke, 2005). Informational masking in phonetic recognition mainly depends on the interference on acoustic-phonetic cues of speech sounds, e.g., the informational masking at the phonetic level may primarily come from spectral misallocation.

2.1.3 Language experience: native and nonnative listeners

Learning the phonetic system of a second language is difficult for non-native listeners. Previous studies found that nonnative listeners usually have disadvantages in using acoustic-phonetic cues on phonetic perception (Cutler, Smits, & Cooper, 2005; Flege et al., 1997; Kondaurova & Francis, 2008; Liu et al., 2012; Mi et al., 2016;

Morrison, 2009; Tyler & Cutler, 2009; Wang, 2006; Ylinen et al., 2010). For example, nonnative listeners tend to employ perceptual weighting on acoustic-phonetic cues differently from native listeners. Formant and duration are two acoustic-phonetic cues in vowel identification. Native listeners primarily depend on F1 and F2 to recognize vowels (Hillenbrand et al., 2001; Peterson & Barney, 1951; Strange, 1989). The effect of the duration on vowel identification is usually limited (Hillenbrand, Getty, Clark, & Wheeler, 1995). On the other hand, nonnative listeners showed a higher perceptual weight of duration and lower weight of formant cues compared with native listeners (Bohn, 1995; Cebrian, 2006; Escudero, Benders, & Lipski, 2009; Giannakopoulou, 2012; Hsieh & Pan, 2010; Lipski, Escudero, & Benders, 2012; Morrison, 2009; Munro, 1993; Wang, 2006). Munro (1993) found that the Arabic-native listeners relied on the duration cue to distinguish English vowels /i/ and /ɪ/. Similar conclusions were also reported for nonnative listeners with other language backgrounds, e.g., Japanese, Spanish, Greek and Mandarin Chinese (Bohn, 1995; Giannakopoulou, 2012; Hsieh & Pan, 2010; Morrison, 2009; Wang, 2006), and for distinguishing other English vowel pairs (e.g., /u/-/ʊ/ and /æ/-/ɛ/; heed - hid and who'd – hood) (Hsieh & Pan, 2010; Wang, 2006). For the task of English vowel formant discrimination, Liu et al. (2012) found that the non-native listeners had higher thresholds than native listeners. These studies showed nonnative listeners usually have a poorer capacity to use formant cues for speech recognition, and they may depend more extensively on the duration cue on vowel identification. In addition, language learning experience and perception training would also affect the

weight of spectral and duration cues. Compared with inexperienced nonnative listeners, experienced nonnative listeners usually depend more on formant cues instead of the duration cue (Flege et al., 1997; Hsieh & Pan, 2010). Ylinen et al. (2010) found that perception training could increase formant cues' perceptual weight on nonnative listeners' vowel identification.

In general, nonnative listeners suffer more interferences from background noise than native listeners. In the study by Gat and Keith (1978), English native and nonnative listeners were required to take a test of English word recognition in quiet and white noise with three SNRs (e.g., 0, 6 and 12 dB). Results suggested that native and nonnative listeners showed comparable performances in quiet. However, compared with native listeners, the recognition scores of nonnative listeners dropped more with the decrease of SNR in noise conditions. Several follow-up studies also demonstrated this conclusion (Crandell & Smaldino, 1996; Rosenhouse, Haik, & Kishon-Rabin, 2006; Stuart, Zhang, & Swink, 2010). Jin and Liu (2012) focused on the nonnative disadvantages at low SNRs. In their study, native (e.g., English-native) and nonnative (e.g., Mandarin Chinese-native and Korean-native) listeners were recruited for sentence recognition tests in quiet, LTSSN and MTB at different SNRs (e.g., -10 – 5 dB for LTSS noise and -15 - 0 dB for MTB noise). As a result, nonnative listeners suffered more interferences from background noise, especially at moderate SNRs (e.g., -5 dB SNR for LTSS noise and -10 dB SNR for babble).

The nonnative disadvantages in phonetic perception in noise is different from those in sentence recognition. For vowel identification, (Cutler, Weber, Smits, & Cooper, 2004) found that nonnative (e.g., Dutch-native) listeners showed a poorer performance in English vowel identification in both quiet and MTB at high SNRs (e.g., 0, 8, 16 dB), and the nonnative disadvantages were not enlarged in noise conditions. Another study by Cutler et al. (2005) also drew a similar conclusion. These studies showed that nonnative listeners did not suffer more interferences from background noise at high SNRs (e.g., from 16 dB to 0 dB) in vowel identification than native listeners. Mi et al. (2013) investigated vowel identification for native and nonnative listeners (e.g., Chinese-native listeners in China) in adverse LTSSN and MTB at SNRs -15 dB to 0 dB. Results suggested that the nonnative disadvantages in vowel identification decreased in LTSSN but increased in MTB at SNRs of -3 dB and -6 dB. These results indicated that the nonnative disadvantage in vowel identification in noise depends on noise type, SNR, and nonnative listeners' language experience. For consonant identification, Nábělek and Donahue (1984) compared identification scores for native (English-native) and nonnative listeners with various native languages (e.g., Polish, Spanish, Japanese, Hungarian and Chinese) in listening conditions of quiet and reverberation. Nonnative listeners performed comparably well with native listeners in quiet but suffered more difficulties in the presence of reverberations. Similar results were also found by Takata and Nábělek (1990) for nonnative listeners with Japanese as the first language. In addition, Lecumberri and Cooke (2006) investigated English consonant identification for non-native listeners in

various listening conditions, e.g., in quiet, LTSSN, MTB and competing speech noise at the SNR of 0 dB. Results showed that nonnative disadvantages in consonant identification were enlarged in noise conditions compared with that in quiet, suggesting that nonnative listeners suffered more from background noises at moderate SNRs than native peers. Tao et al. (2018) explored English consonant perception for Mandarin Chinese-native listeners with SNRs from -20 dB to 0 dB. They found more significant nonnative disadvantages for Chinese-native listeners in China in multi-talker babble at low SNRs (e.g., from -20 dB to -10 dB) than in quiet.

The cross-linguistic studies above showed that nonnative listeners suffered more in dynamic speech noise (e.g., MTB) compared with stationary nonspeech noise (e.g., LTSSN), which occurred in both sentence recognition (Jin & Liu, 2012) and phonetic identification (Cooke et al., 2008b; Mi et al., 2013; Tao et al., 2018). Two possible mechanisms were proposed by Mi et al. (2013). One is that nonnative listeners might receive less masking release from temporal glimpses in background noise. Compared with LTSSN, listeners can take advantage of temporal dips in babble to improve speech perception. With English-native and Mandarin Chinese-native participants as native and nonnative listeners, Stuart et al. (2010) found more nonnative disadvantages in interrupted noise compared with stationary noise in a task of English sentence recognition, indicating that Chinese-native listeners had a lower masking release from temporal dip listening compared to English-native listeners. Guan et al. (2015) confirmed the finding that Chinese-native listeners took less advantage in temporal dip listening

than English-native listeners in vowel identification in temporal-fluctuating noise. In addition, several studies found that native English exposure could improve temporal dip listening for nonnative listeners (Guan et al., 2015; Li et al., 2016; Mi et al., 2013). Another possibility was that, compared with native listeners, nonnative listeners might be affected more by informational masking in MTB. The informational masking is affected by the familiarity of speech information in babble noise (Van Engen, 2010). When the target speech and background noise are in the same language, nonnative listeners usually suffer more informational masking due to their less familiarity with MTB, possibly resulting in their greater difficulty separating speech signals from the background babble noise.

2.2 Strategies to improve speech perception in noise

Adverse background noise reduces the clarity and intelligibility of speech by degrading acoustic-phonetic cues (e.g., formants) of speech, usually leading to more difficulties in speech perception. To improve speech recognition in noise, two major strategies in signal processing are generally used: noise reduction and speech enhancement. Noise reduction aims to reduce background noise intensity based on the separation of target speech and background noise. In contrast, speech enhancement focuses on strengthening acoustic cues of target speech sounds.

2.2.1 Noise reduction

The first and critical step of noise reduction is to distinguish speech signals from background noise. Directional microphones and noise reduction algorithms are two effective methods of noise reduction (Chung, 2004). An array of directional microphones is used to detect spatial differences This technology assumes assumption underlying this technology is that listeners in a complex environment (e.g., a restaurant) usually orient toward each signal of interest in turn (Brimijoin, Whitmer, McShefferty, & Akeroyd, 2014), e.g., if target speech comes from the front of listeners, it will reach the front microphone first then followed by the rear microphone. Directional microphones effectively separate target speech and background noise, and this technique has been widely used in hearing devices. However, this technique also has some limitations, e.g., it depends on the hearing devices with directional microphones. In addition, the effect is limited when target speech and background noise come from the same direction.

Another noise reduction strategy aims to distinguish target speech and background noise based on the temporal or spectral characteristics (for a review, see Bentler & Chiou, 2006). The incoming sounds with these temporal (e.g., modulations with rate around 3 Hz and modulation depths at midfrequency ranges to approximately 30 dB - 50 dB) and spectral features are considered as the target signals with others as noise (Plomp, 2019). In noise reduction, once target speech and background noise are recognized, the target speech would be amplified with the control on the gains for background noise. In addition, noise reduction algorithms are available for hearing devices without an array of directional microphones and for speech combined with noise

from the same direction. However, it is also challenging to distinguish target speech from background noise mixed intractably (Chung, 2004).

2.2.2 Speech enhancement

Speech enhancement is another technical approach to improve speech perception in noise. Speech enhancement aims to restore or strengthen the acoustic cues of target speech masked or smeared by background noise. As one type of speech enhancement, spectral enhancement was initially designed in amplification systems for hearing-impaired listeners. Some studies suggested this rationale might help listeners with normal hearing (Guan & Liu, 2019a; Guan & Liu, 2019b; Lyzenga, Festen, & Houtgast, 2002). The benefits of spectral enhancement depend on the enhancement strategies and specific techniques.

One traditional strategy is to enhance spectral cues of target speech mixed with background noise. Two major solutions to restore smeared spectral cues with background noise are spectral sharpening and spectral expansion. Spectral sharpening aims to strengthen spectral cues by narrowing formant bandwidth (Summerfield et al., 1985), while spectral expansion enhances spectral contrasts by amplifying peaks and/or attenuating valleys with weighted spectral filters (Boers, 1980; Bunnell, 1990; Chen & Wang, 2011; Franck et al., 1999a; Lyzenga et al., 2002; Stone & Moore, 1992).

Summerfield et al. (1985) investigated the effect of spectral bandwidth on consonant identification for listeners with normal hearing and hearing loss. Listeners

were asked to recognize synthetic /CVC/ (i.e., consonant-vowel-consonant) syllables the with normal, broadened and narrowed spectral bandwidth in quiet conditions. Results suggested that the broadened bandwidth reduced identification accuracy for listeners with normal hearing and hearing loss; however, narrowed bandwidth did not significantly increase the accuracy. Thus, although spectral sharpening did not improve speech perception in quiet, it does not mean spectral sharpening has no effect on speech perception in noise. In the study by Lyzenga et al. (2002), the expansion of spectral contrast alone was not enough to facilitate speech perception in noise. However, after combining spectral expansion and sharpening, sentence reception was significantly improved in LTSSN for about 1 dB. These results suggested that combination of spectral sharpening and expansion might be helpful to improve speech perception in noise for normal-hearing listeners.

Bunnell (1990) investigated the effects of spectral expansion on consonant identification in quiet for listeners with moderate to severe sensorineural hearing loss. A weighted inverse filter was applied to the spectral envelope that was estimated over consecutive and overlapping (50%) 25.6-ms speech segments. The enhancement resulted in approximately 5% improvements in consonant identification. Simpson et al. (1990) explored the effect of spectral enhancement with background noise. Different from the approach in the study by Bunnell (1990), an adaptive filter bank was designed to model the filtering properties and auditory excitation patterns of normal ears. Then, the excitation pattern was enhanced by the convolution with a function of difference-of-

Gaussians (DOG) or "Mexican Hat," i.e., amplification of the spectral peaks and attenuation of the adjacent valleys in the excitation pattern. With this algorithm of spectral enhancement, listeners with hearing loss gained small but significant improvements in sentence recognition in LTSSN. To speed up signal processing, Stone and Moore (1992) applied an analog filter corresponding to the method in the study by Simpson et al. (1990). A 16-channel analog filter bank was applied based on a logarithmic audio-frequency scale with bandwidths approximating those of auditory filter. The channels with local maxima were amplified as spectral peaks, while the neighboring channels were attenuated as spectral valleys. Although spectral contrasts were effectively increased with the analog enhancement, hearing-impaired listeners could not gain significant improvement of speech perception in noise.

Baer et al. (1993) developed the technique in Simpson et al. (1990) with the fast-acting amplitude compressions for enhanced gains and compared the perceptual effects of the enhancements with and without compression. They found that the enhancement with compression resulted in more significant on speech intelligibility (e.g., equivalent to a 4.2 dB increase of the SNR) and shorter response time for hearing-impaired listeners than the algorithm without compression. Based on the modified enhancement in the study by Baer et al. (1993), Yang et al. (2003) further developed a simulation model to calculate the changes in acoustic features and internal representation (e.g., excitation pattern) of speech signals with spectral enhancement. The new enhancement algorithm achieved desired increase of spectral contrasts with an appropriate selection of related parameters (e.g.,

different bandwidth and enhancement scale in several types of noise at various SNRs). However, there was a lack of behavioral experiments to examine the enhancement effect on speech perception for human listeners. Rout (2006) considered the factor of the frequency range for spectral enhancement. In his study, the frequency range for spectral enhancement was widened from 5000 Hz in previous studies (Bunnell, 1990; Simpson et al., 1990) to 8000 Hz with a similar approach. However, this enhancement algorithm did not result in any perceptual improvements, i.e., it did not improve sentence recognition for listeners with hearing loss, and listeners preferred the quality of unmodified sounds without enhancement. Considering the channel number of compressions, Franck et al. (1999) compared the enhancement algorithms between single and multi-channel amplitude compressions. They reported that the enhancement with multichannel compression significantly improved vowel identification for hearing-impaired listeners in LTSSN, while the single compressor did not improve substantially. However, the perceptual benefit was found only in vowel identification but not in consonant recognition.

As speech information is carried in dynamic spectral changes instead of static shapes, Chen et al. (2012) developed an enhancement algorithm with a dynamic spectral scheme. Unlike the static enhancement in previous studies, the new algorithm dynamically enhanced spectral cues over time based on the overlap-add method. In addition, the effects of the dynamic algorithm on speech intelligibility were tested for hearing-impaired listeners in several noise conditions, e.g., LTSSN and two-talker babble

with the SNRs of -6 and -3 dB. As a result, a small but significant benefit (e.g., about 8%) from the dynamic enhancement was found in speech intelligibility in LTSS noise at -6 dB SNR, but not in other noise conditions (e.g., LTSSN at -3 dB SNR and MTB at -6 dB and -3 dB SNR). In follow-up research, as the enhancement parameters were customized for each participant, the intelligibility improvement from the dynamic enhancement could reach 14% (J. Chen, Baer, & Moore, 2013). However, the improvement was only observed in LTSS noise at -6 dB SNR and was not found at other noise conditions.

In summary, most of the previous studies suggested that spectral enhancement had small (3-5%), or even no benefit on speech perception in noise (Bunnell, 1990; Franck et al., 1999a; Rout, 2006; Stone & Moore, 1992; Summerfield et al., 1985). In the study by Chen et al. (2013), although the benefits on speech intelligibility could reach 8%, and even 14% after individualizing each participant' enhancement parameters, the perceptual benefits from spectral enhancement were limited to certain noise conditions, e.g., in LTSSN at -6 dB SNR. There are two possible reasons for the limitations of these enhancement algorithms. First, the enhancement techniques were applied to target speech and backgrounds simultaneously. That is, both target speech and background noise in specific spectral ranges are enhanced together. However, the local SNRs of spectral peaks, which speech perception in noise primarily depends on, might not be improved significantly. Second, as these techniques enhance spectral peaks across a broad frequency range, the spectral prominences at low frequencies may have masking effects

on high-frequency peaks, i.e., the upward spread of masking (Egan & Hake, 1950; Wegel & Lane, 1924). Compared with low-frequency cues (e.g., F1), the spectral cues at mid and high frequencies (e.g., F2 and F3) generally have lower intensities, which are less reliably detected in challenging listening conditions. Therefore, the upward spread of masking with low frequency amplification might offset the benefits from spectral enhancement at high frequencies.

In addition to the spectral enhancement described above, a new enhancement rationale, named Contrast Enhanced Frequency Shaping (CEFS), was proposed based on the representation of spectral contrasts in the peripheral auditory system (Bruce, 2004; Miller, Calhoun, & Young, 1999). Compared with F1 at low frequencies, F2 and F3 at mid and high frequencies are more susceptible to spectral smearing with background noise, especially for listeners with high-frequency hearing loss. Therefore, CEFS argued that an effective enhancement algorithm should focus on the amplifying of spectral cues at high frequencies (e.g., F2 and F3) to address spectral smearing at high frequency. By recording neural responses of cats with noise-induced hearing loss, Miller et al. (1999) found that the CEFS-enhanced vowels (i.e., spectral enhancement of F2 and F3) resulted in better representation of F1 and F2, as well as suppression of neural fibers for spectral valleys. Bruce (2004) developed the CEFS enhancement with multichannel compressions and suggested no distortion of formant representation with the computation models of normal and impaired auditory neurons. The high-frequency amplification in CEFS enhancement inspired for the enhancement algorithm in the dissertation study, which

focused on the amplification of the high-frequency cues (e.g., F2) without disrupting low frequency information (e.g., F1).

2.2.3 F2 enhancement in the current study

A new relatively straightforward algorithm of spectral enhancement, called F2 enhancement, was recently proposed in our laboratory (Guan & Liu, 2019a; Guan & Liu, 2019b; Woodall & Liu, 2013). This F2 enhancement is directly applied to target speech without changing background noise, resulting in greater local SNRs of formants. There are four steps in the F2 enhancement procedure, e.g., (1) render a 3-D spectrogram (amplitude × time × frequency) with Fast Fourier transform (FFT); (2) estimation of F2 based on formant peak and surrounding spectral valleys; (3) enhancement of the F2 peak; and (4) resynthesize F2-enhanced speech with an inverse FFT.

Woodall and Liu (2013) investigated the effects of F2 enhancement with various enhancement scales (e.g., 3, 6, and 9 dB) on vowel formant discrimination in quiet for listeners with normal hearing and sensorineural hearing loss. Thresholds of vowel formant discrimination refer to the smallest changes in formant frequency that is detectable. Results suggested that F2 enhancement significantly improved hearing-impaired listeners' sensitivity to vowel formant frequency change, e.g., 46% improvements (e.g., reduction of formant discrimination thresholds) from F2 enhancement at 3 dB, 60% at 6 dB, and 71% at 9 dB. However, the improvement in vowel formant sensitivity for hearing-impaired listeners was not found for listeners with

normal hearing, possibly because that the spectral contrast of F2 for the unmodified signals was large enough. Guan and Liu (2019b) further explored the effect of F2 enhancement on formant discrimination in LTSS noise for older listeners with normal and impaired hearing. F2 enhancement at 9 dB significantly improved vowel formant sensitivity in LTSSN for both groups of older listeners. In addition, the perceptual improvement at SNR of 6 dB was higher than SNR at 12 dB. In another study, Guan and Liu (2019a) measured speech recognition thresholds of coordinate response corpus (CRM) in MTBs with and without F2 enhancement for English and Chinese speech. F2 enhancement significantly improved speech recognition in two-talker babble, but in six-talker babble for both languages. In summary, the previous studies in our laboratory found that F2 enhancement could lead to perceptual improvements in vowel formant discrimination and word recognition, depending on listening conditions.

Although F2 enhancement had indicated perceptual benefits in speech perception in quiet and noise, there were still puzzles of the perceptual effects of F2 enhancement. First, as the previous studies of F2 enhancement concentrated on vowel formant discrimination and word recognition, it was unknown whether the significant improvements would occur in phonetic identification, e.g., in vowel identification, consonant identification, or both. In particular, one research question was the improvement in word recognition by F2 enhancement (Guan and Liu, 2019a) was due to the perceptual benefits from vowel perception, consonant perception, or both. Second, dynamic speech noises (e.g., MTB) were used as the maskers in the study by Guan and

Liu (2019a); meanwhile, it was unclear if the improvements would occur in stationary nonspeech noise without temporal dips and speech information. In other words, another research question was whether listeners could benefit from the task of speech perception in fixed noise that is primarily composed of energetic masking, by F2 enhancement. Third, given a significant improvement in speech perception for native listeners with normal and impaired hearing (Guan & Liu, 2019a; Guan & Liu, 2019b; Woodall & Liu, 2013), the third research question was whether the benefit of F2 enhancement could be found for nonnative listeners who face significant challenges of speech recognition in noise.

2.3 Goals of the current study

First, the current study aimed to investigate whether F2 enhancement could improve phonetic identification with the factors of phonetic type, noise type, SNR, and listeners' language experience, e.g., vowel and consonant identification for native and nonnative listeners in various noise conditions of speech and nonspeech noise at moderately and very challenging SNRs. Second, this study examined the amount of perceptual benefit from F2 enhancement, if any, would be comparable or vary on these factors (e.g., vowel identification vs consonant identification). The hypotheses included: (1) the F2 enhancement would improve vowel and consonant identification in noise, given the F2 importance in both vowel and consonant processing. In addition, greater enhanced benefits were expected on vowel identification due to its higher perceptual

weighting of F2 compared with consonant identification, e.g., F2 is a primary but less reliably detected cue in vowel identification (Parikh & Loizou, 2005); meanwhile, F2 transition is only one of the phonetic cues in consonant perception (e.g., burst release, F1 and F3 transitions) (Dorman et al., 1977; Kewley-Port, 1982; Story & Bunton, 2010). (2). Since F2 peaks of speech signals are usually smeared because of noise masking, significant improvements from F2 enhancement were expected for both types of noise: LTSSN and MTB. Besides, greater benefits were expected in speech noise than in non-speech noise, and at low SNRs than at high SNRs due to more elevated noise masking in these conditions. (3) As both native and nonnative listeners depend on F2 information on phonetic identification, both groups would gain benefits from the enhanced F2. Moreover, it was expected that nonnative listeners would gain more improvements from F2 enhancement than native listeners. Due to their disadvantages in formant processing and capacities against noise masking, nonnative listeners usually require higher audibility of speech formants in challenging listening conditions compared with native listeners. Thus, enhanced F2 information might be more beneficial for nonnative listeners in phoneme identification in noise than native listeners.

Chapter 3: Methods

3.1 Participants

Two groups of participants, 16 young English-native (EN) and 20 Mandarin Chinese-native (CN) listeners were recruited in this study. The EN listeners were recruited at the University of Texas at Austin (UT), and the CN listeners were recruited at Shanghai Jiao Tong University (SJTU). The demographic characteristics of the participants were listed in Table 1. The listeners in two groups were matched in age (t = 1.76, $p > 0.05$) and gender ($\chi^2 = 3.6$, $p > 0.05$). All the listeners have pure tone thresholds less or equal to 20 dB HL at octave intervals between 250 and 8000 Hz (ANSI, 2010). CN listeners started their school-based English education from 6 to 13 years old. The CN listeners were asked to complete the Questionnaire for Bilingual Speakers (see Appendix 1) to collect the information of their English learning experience (e.g., age of English acquisition, length of English learning and usage ratio of English-to-Chinese, and self-evaluation of their English proficiency). Table 2 shows that CN listeners had a long length of English learning (e.g., 13.61 years) but limited usage ratio of English-to-Chinese (e.g., 0.16) in average. As shown in Table 3, the overall English proficiency for CN listeners was at an average level, i.e., they could understand and speak English with some difficulties.

Table 1: Demographic characteristics of English-native (EN) and Chinese-native (CN) listeners

	EN listeners	CN listeners
Number of participants	16	20
Age range	19 - 23	19 - 27
Mean age (STD)	20.69 (1.25)	21.95 (2.51)
Male/female	5/11	10/10

Table 2: English learning information for CN listeners

	Mean	Standard deviation
Age of English acquisition (year)	8.34	0.53
Length of English learning (year)	13.61	2.81
Usage ratio of English-to-Chinese	0.16	0.16

Table 3: Self-evaluation for English proficiency for CN listeners

	Mean	Standard deviation
Overall proficiency	3	0.69
Reading	3.5	0.79
Speaking	2.33	0.84
Listening	2.78	0.81
Writing	2.83	0.86

3.2 Speech stimuli and noise

There were two experimental tasks, e.g., vowel and consonant identification, in the current study. Two types of stimuli, e.g., unmodified and enhanced speech, were used in each task. Five English front vowels /æ, e, ɛ, i, ɪ/ in the syllabic context of /hVd/ (e.g., had, hayed, head, heed, hid) served as speech stimuli in vowel identification with the consideration of controlling the effect of dialect (e.g., significant /ɔ/-/ɑ/ confusion in Texans). In these vowels, Mandarin Chinese and American English share /e, i/ in common, while English vowels /æ, ɛ, ɪ/ are considered as the "unfamiliar" or "new" for Chinese-native listeners without a counterpart in Mandarin Chinese (Luo, 2002). In the task of consonant identification, six American English stops /p, b, t, d, k, g/ in /aCa/ syllables (e.g., apa, aba, ata, ada, aka, aga) were selected. All these English stop consonants have counterparts in Mandarin Chinese (Duanmu, 2007; Wang, 2007). All

unmodified speech stimuli were recorded from a young female American English native speaker.

The unmodified speech served as the base stimuli for F2 enhancement. Therefore, the F2 enhancement was manipulated for the syllables of the unmodified speech (e.g., the unmodified and F2-enhanced /æ/ shown in Figure 3). The procedure of F2 enhancement was briefly described as four steps as follows, consistent with previous studies (Guan & Liu, 2019a; 2019b; Woodall & Liu, 2013).

The first step was to gain the acoustic information of amplitude × time × frequency of unmodified speech. A three-dimensional spectrogram was created in a MATLAB program for each syllable as a stimulus with the sampling rate at 44,100 Hz, and the window size a 10 ms. A 50% overlap was conducted between the short windows.

Second, speech stimuli' formant peaks and spectral valleys of speech stimuli were located at each 10-ms window, e.g., the frequency with the highest amplitude in a specific frequency range of vowel formant was considered as a formant peak. In contrast, the lowest amplitude between formant peaks was taken as the valley. The formant regions were defined with the peaks and valleys, e.g., F2 area of the second valley (V2) –second peak (P2) – third valley (V3).

The third step aimed to amplify the intensity of F2 peaks. The level of the F2 peak was amplified by 12 dB in this study, and the surrounding components in the F2 region were proportionally enhanced at each time frame. If no F2 peak is located (e.g., consonant segments), no amplification would be applied in the time window.

The last step was to resynthesize the stimuli with F2 enhancement. An inverse FFT was implemented to convert the acoustic information of enhanced speech into audio files.

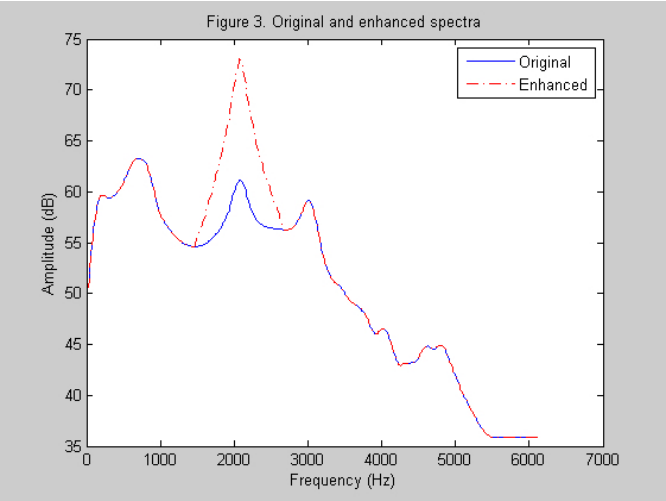

Figure 3. The unmodified (solid blue line) and enhanced (dashed red line) spectra of vowel /æ/

Acoustic analyses were conducted to measure the enhanced scale of the output stimuli. First, the overall intensity of unmodified and enhanced speech was normalized to a fixed level (e.g., 60 dB SPL). Second, the spectra with linear predictive coding (LPC) analyses were conducted for unmodified and enhanced speech after the intensity normalization. Third, F2 intensities of unmodified and enhanced speech were measured and the enhanced scale was computed by subtracting the F2 intensity of unmodified

speech from that of enhanced speech. The actual enhancement of F2 in /hVd/ and /aCa/ are illustrated in Tables 4 and5.

Table 4: The actual F2 enhanced scales in vowel identification

Stimuli	Actual enhanced scale (dB)
Had	6.92
Hayed	7.1
Head	5.79
Heed	8.69
Hid	8.51

Table 5: The actual F2 enhanced scales in consonant identification

Stimuli	Actual enhanced scale (dB)
apa	7.68
aba	7.45
ata	8.08
ada	8.97
aka	8.85
aga	8.47

Two types of noise, e.g., LTSSN and 6-TB, served as maskers in the current study. The 6-TB was generated by recording six (three males and three females) native English adult talkers reading three paragraphs in the technology section in The New Children's Encyclopedia (Lock, 2009), equalizing the intensity of all speech recordings, and then mixing the level-equalized speech recordings of six talkers. The babble-modulated noise was generated by multiplying the temporal envelope of the babble waveform on LTSS noise, which was generated by shaping the Gaussian noise with the filter of the 6-TB average spectrum. The SNRs were manipulated at -10 dB and -15 dB with the noise level fixed at 70 dB SPL and speech level at 60 dB and 55 dB SPL, respectively. No enhancement processing was conducted for background noise.

3.3 Stimulus presentation

For each trial of phonetic identification, a one-sec masker was randomly selected from a 30-sec long six-talker babble or LTSSN. Speech signals were played temporally in the middle of the 1-sec masker. Both speech and masker had a rise/fall time of 10 ms.

Speech and noise were presented at a sampling rate of 24,414 Hz to listeners' right ears via SONY MDR-7506 headphones. The stimulus presentation was controlled by TDT modules, including a two-channel, 24-bit, real-time processor (RP2.1) and a headphone buffer (TDT HB7) at UT, and a mobile sound processor (RM1) at SJTU. For calibration purposes, target speech was normalized to the same root-mean-square (RMS) level. In addition, stimulus and noise levels were calibrated with an AEC201-A IEC 60318-1 ear simulator by a Larson-Davis sound-level meter (Model 2800) with a linear weighting band.

3.4 Procedure

EN listeners conducted the experiment in a sound-treated booth in the Speech Psychophysical Laboratory at UT. In contrast, CN listeners' data were collected in a quiet test room in the Psycholinguistic Laboratory at SJTU.

There were two experimental tasks, e.g., vowel and consonant identification, in the current study. Listeners in each group were balanced for the sequence, e.g., half of them did the vowel identification first, while the other half began with consonant identification. There were five response alternatives in vowel identification and six consonant identification presented in a text box corresponding with each stimulus.

Listeners were seated in front of an LCD monitor and required to click a computer mouse on the button corresponding with their response choice within 10s after each stimulus presentation.

Prior to the data collection of each task, listeners had a 5-mins practice session of unmodified vowel/consonant identification quiet was conducted to have participants get familiar with the experimental procedure. Feedback was provided to indicate the correct response on each trial in practice session, while no feedback was provided during the formal tasks. In each formal task, phonetic identification without and with F2 enhancement was conducted in the listening conditions of quiet, LTSSN and 6-TB at the SNRs of -15 dB and -10 dB. Listeners took the formal identification task in the quiet condition first, followed by with the noise conditions with a mixed and random order. In addition, the conditions of unmodified and enhanced stimuli were also randomized in quiet and noise conditions. Under each listening condition, each stimulus was presented for 15 times and all stimuli (e.g., five stimuli in vowel identification and six stimuli in consonant identification) were presented randomly; thus, for each condition, vowel identification in percent-correctness was computed on the 15 judgments for each vowel. There were 20 conditions in total (e.g., unmodified, and enhanced speech in the conditions of quiet, LTSS and 6-TB at the SNRs of -15 and -10 dB in a vowel and consonant identification), and each condition took about 4-6 mins. Training and formal experiments were completed for approximately 2 hours, and short breaks were provided between blocks. The software SYKOFIZX was used to implement the procedure.

Chapter 4: Results

The correctness percentages (i.e., percentage of identification accuracy) for phonetic identification were used as the dependent variables in statistical analyses. The multiway analyses of variance (ANOVAs) were applied to examine the significance of the main factors and interaction effects. Additionally, Bonferroni correction was used for the *post hoc* comparisons for significant main effects and the simple effect analysis for significant interaction effects.

4.1 Vowel identification

4.1.1 Identification of unmodified vowels

4.1.1.1 Identification of unmodified vowels in quiet

The identification percentages of the five English vowels (e.g., /æ, e, ɛ, i, ɪ/) for EN and CN listeners were shown in Figure 4. A two-way (within-subjects factor: vowel category; between-subjects factor: listener group) ANOVA was conducted for the unmodified vowel identification for EN and CN listeners. The main effects of listener group ($F_{1, 32} = 36$, $p < 0.01$, $\eta_p^2 = 0.53$) and vowel category were significant ($F_{4, 128} = 7.49$, $p < 0.01$, $\eta_p^2 = 0.19$). In addition, although the interaction effect of listener group × vowel category was significant ($F_{4, 128} = 6.79$, $p < 0.01$, $\eta_p^2 = 0.18$), the simple effect analysis compared by listener group suggested that EN listeners performed significantly better than CN listeners in the identification of all five English unmodified vowels (all *ps* < 0.01; shown in Figure 4).

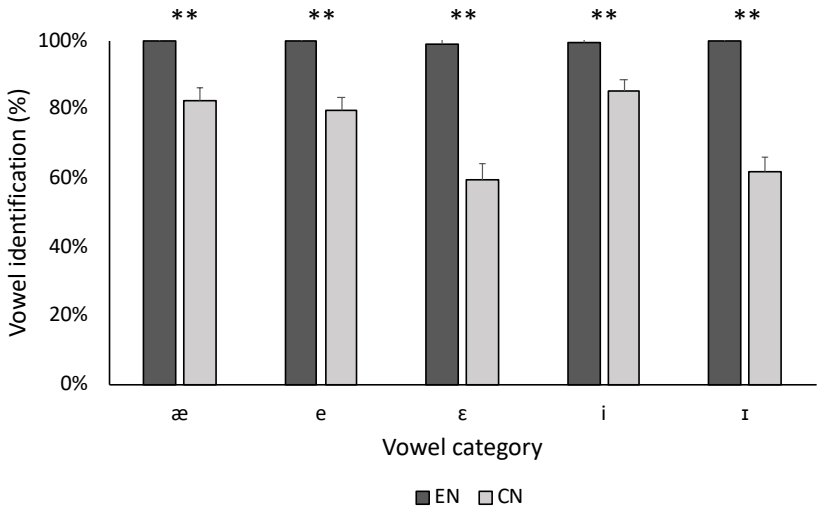

Figure 4. Identification of unmodified vowels (/æ, e, ɛ, i, ɪ/) for English-native (EN) and Chinese-native (CN) listeners in quiet.

4.1.1.2 Identification of unmodified vowels in noise

Figure 5 illustrated average percentages of unmodified vowel identification in LTSSN and 6-TB at the SNRs of -15 dB and -10 dB for EN and CN listeners. A four-way ANOVA (within-subjects factors: SNR, noise type and vowel category; between-subjects factor: listener group) was conducted, and the results revealed significant main effects of SNR ($F_{1, 34} = 144.41$, $p < 0.01$, $\eta_p^2 = 0.81$), noise type ($F_{1, 34} = 14.97$, $p < 0.01$, $\eta_p^2 = 0.31$), vowel category ($F_{4, 128} = 52.21$, $p < 0.01$, $\eta_p^2 = 0.61$), as well as listener group ($F_{1, 34} = 27.54$, $p < 0.01$, $\eta_p^2 = 0.45$). The *post hoc* tests of these effects suggested lower vowel

identification at SNR of -15 dB than -10 dB, in 6-TB than in LTSSN, for non-native listeners than native listeners (all *ps* < 0.01).

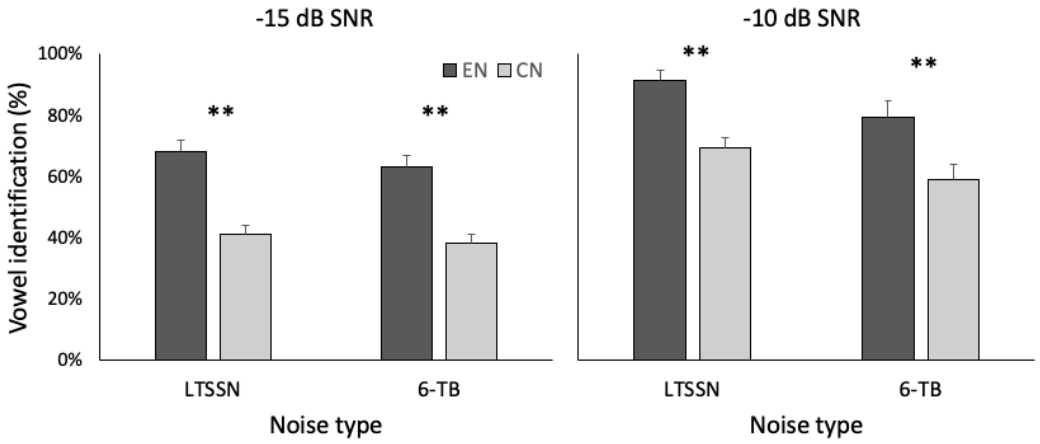

Figure 5. Identification of vowels in LTSSN and 6-TB at the SNRs of -15 dB (left) and -10 dB (right) for EN and CN listeners. Error bars indicate standard error. ** $p < 0.01$.

The significant interaction effects were listed in the Table 6. The simple effect analysis of SNR × noise type suggested vowel identification in 6-TB was lower than in LTSSN when the SNR was at -10 dB ($p < 0.01$), while the identification score was comparable in the two types of noise at the SNR of -15 dB ($p > 0.05$). In addition, the significant interaction effect of listener group × SNR × vowel category showed better

identification for EN listeners than for CN listeners at most conditions (all *ps* < 0.05) except for /i/ at the SNRs of both -15 dB and -10 dB (both *ps* > 0.05).

Table 6: Significant interaction effects on vowel identification in noise

	Interaction effect	*F*	*p*	η_p^2
Two-factor	Listener group × vowel category	4.03	0.004	0.11
	SNR × noise type	6.48	0.016	0.16
	SNR × vowel category	19.04	0.000	0.36
Three-factor	Listener group × SNR × vowel category	6.91	0.000	0.17

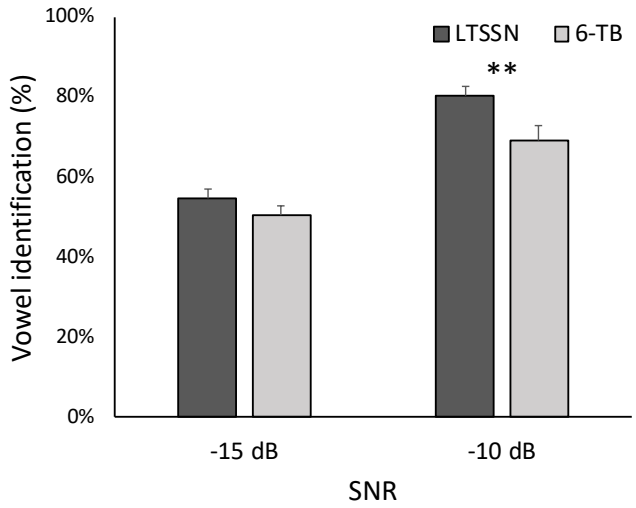

Figure 6. Identification of unmodified vowels in LTSSN and 6-TB at the SNRs of -15 dB and -10 dB. Error bars indicate standard error. **p < 0.01.

4.1.2 The effect of F2 enhancement on vowel identification in quiet and noise

4.1.2.1 The effect of F2 enhancement on vowel identification in quiet

Figure 7 showed the average percentages of the vowel identification with and without F2 enhancement for EN and CN listeners. A three-way (within-subjects factors: enhancement and vowel category; between-subjects factor: listener group) ANOVA was conducted to explore the effect of F2 enhancement in quiet conditions. The main effect of enhancement was not significant ($F_{1, 34} = 0.474$, $p > 0.05$, $\eta_p^2 = 0.02$), suggesting that the F2 enhancement did not significantly affect vowel identification in quiet. In addition, there were no significant interaction effects of either listener group or vowel category with the factor of enhancement (both $p > 0.05$).

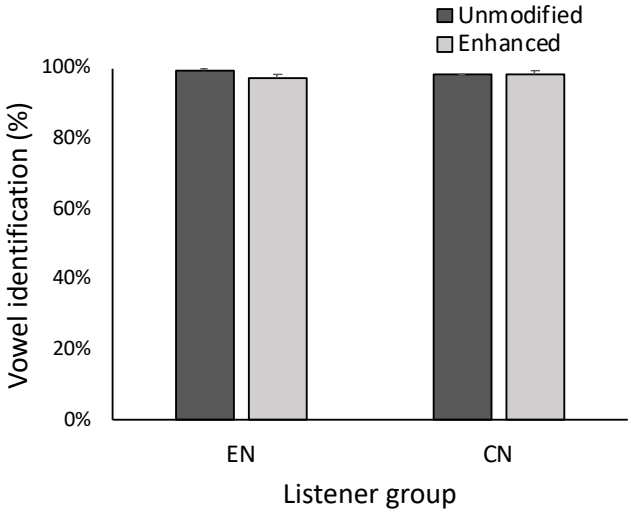

Figure 7. Identification of unmodified and enhanced vowels for EN and CN listeners in quiet. Error bars indicate standard error.

4.1.2.2 The effect of F2 enhancement on vowel identification in noise

A five-way (within-subjects factors: enhancement, SNR, noise type and vowel category; between-subjects factor: listener group) ANOVA was conducted to investigate the effects of F2 enhancement in noise conditions. Results suggested main effects of all the five factors were significant (all *ps* < 0.01). The significant main effect of enhancement ($F_{1, 34} = 24.37$, $p < 0.01$, $\eta_p^2 = 0.42$) suggested that overall, the identification of enhanced vowels was significantly higher than that of unmodified vowels.

Table 7 listed the significant interaction effects of F2 enhancement and other factors. The simple effect analysis on enhancement × SNR × noise type indicated that the

F2 enhancement improved vowel identification in LTSSN and 6-TB at the SNR of -15 dB and 6-TB at -10 dB SNR (all *ps* < 0.01). However, vowel identification in LTSSN at -10 dB SNR was reduced with the F2 enhancement ($p < 0.01$), as shown in Figure 8. The significant interaction effect of enhancement × vowel category × listener group suggested that the identification scores of /e, i/ (all *ps* < 0.01) were increased with F2 enhancement for both EN and CN listeners. Meanwhile, F2 enhancement negatively affected the identification of /æ, ɛ, ɪ/ for EN listeners (all *ps* < 0.05) and /ɪ/ for CN listeners ($p < 0.05$).

Table 7: Significant interaction effects with the factor of enhancement on vowel identification in noise

	Interaction effect	*F*	*p*	η_p^2
Two-factor	Enhancement × listener group	6.64	0.015	0.16
	Enhancement × SNR	21.09	0.000	0.38
	Enhancement × noise type	12.77	0.001	0.27
	Enhancement × vowel category	116.61	0.000	0.77
Three-factor	Enhancement × SNR × noise type	17.65	0.000	0.34
	Enhancement × vowel category × listener group	3.12	0.017	0.08
	Enhancement × vowel category × SNR	24.05	0.000	0.41
Four-factor	Enhancement × SNR× noise type × vowel category	7.30	0.000	0.18

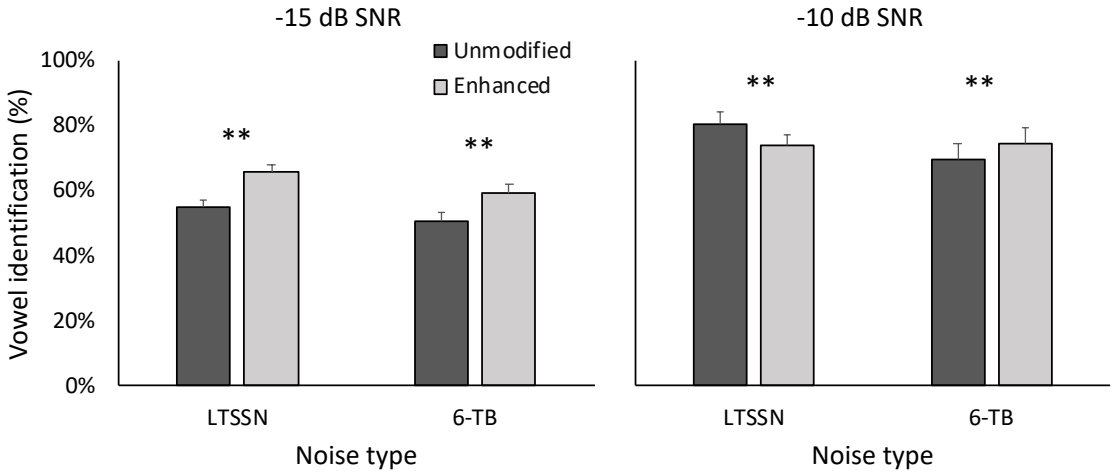

Figure 8. Identification of unmodified and enhanced vowels in LTSSN and 6-TB at the SNRs of -15 dB (left) and -10 dB (right). Error bars indicate standard error. **$p < 0.01$.

4.1.2.3 Effect of F2 enhancement on vowel confusion matrix in noise

Tables 8 and 9 showed the effect of F2 enhancement on vowel confusion matrix in noise for EN and CN listeners, respectively. The effect of F2 enhancement on vowel confusion matrix were defined as the differences of confusion matrices in noise with and without F2 enhancement, which were computed as two steps. The first step was to calculate the confusion matrix of unmodified and enhanced vowels averaged over the noise conditions with two SNRs and noise types. The second one was to subtract the average confusion matrix of unmodified vowels from the average matrix of enhanced

vowels. As a result, the identification percentages of /e, i/ with F2 enhancement in noise were significantly increased for both EN and CN listeners. Meanwhile, the identification scores of /æ, ɛ, ɪ/ were reduced for EN listeners and /ɪ/ was reduced for CN listeners. As shown in Table 8, the confusions of /e, i/ with all other vowels were generally reduced for EN listeners. In addition, the reduced identification of /æ/ (i.e., reduction of 7.3%) and /ɪ/ (i.e., reduction of 18.5%) for EN listeners mainly came from the increased confusions with /ɛ/ (i.e., increases of 6.4% for /æ/ and 29.7% for /ɪ/). The reduced identification of /ɛ/ (i.e., reduction of 5%) was mainly attributed to the confusions increase with /ɪ/ (i.e., increase of 6.3%). For CN listeners, the confusions of vowels /e, i/ with improvement were also reduced similarly with those of EN listeners (see Table 9), while the vowel /ɪ/ with reduced identification (i.e., reduction of 18.5%) might be attributed to increased confusions with both /ɛ/ (i.e., increase of 15.3%) and /æ/ (i.e., increase of 8.7%).

Table 8: Effect of F2 enhancement on vowel confusion matrix for EN listeners

Response / Target	/æ/	/e/	/ɛ/	/i/	/ɪ/
/æ/	**-7.3%**	0.0%	6.4%	0.5%	0.4%
/e/	-1.1%	**11.4%**	-4.0%	-3.4%	-2.9%
/ɛ/	-0.7%	-0.2%	**-5.0%**	-0.4%	6.3%
/i/	-5.4%	-9.8%	-9.0%	**31.8%**	-7.5%
/ɪ/	-0.4%	-4.6%	29.7%	-6.3%	**-18.5%**

Table 9: Effect of F2 enhancement on vowel confusion matrix for CN listeners

Response / Target	/æ/	/e/	/ɛ/	/i/	/ɪ/
/æ/	**-3.7%**	1.1%	2.0%	-1.0%	1.6%
/e/	-4.0%	**11.6%**	-3.8%	-1.0%	-2.8%
/ɛ/	5.3%	-1.8%	**-2.4%**	-3.6%	2.4%
/i/	-7.5%	-10.3%	-8.0%	**34.0%**	-8.1%
/ɪ/	8.7%	-11.2%	15.3%	-8.8%	**-4.0%**

4.2 Consonant identification

4.2.1 Identification of unmodified consonants

4.2.1.1 Identification of unmodified consonants

The percentage correctness of stop consonant identification (e.g., /b, d, g, p, t, k/) for EN and CN listeners was shown in Figure 8. A two-way (within-subjects factor: vowel category; between-subjects factor: listener group) ANOVA was conducted for the unmodified consonant identification for EN and CN listeners in quiet. The results showed significant main effects of listener group ($F_{1, 32} = 6.65$, $p < 0.05$, $\eta_p^2 = 0.16$) and consonant category ($F_{4, 128} = 5.83$, $p < 0.01$, $\eta_p^2 = 0.15$). Overall, EN listeners showed significantly better performance in consonant identification than the CN listeners. In addition, the interaction effect between listener group and vowel category was not significant ($F_{4, 128} = 1.36$, $p > 0.05$, $\eta_p^2 = 0.03$).

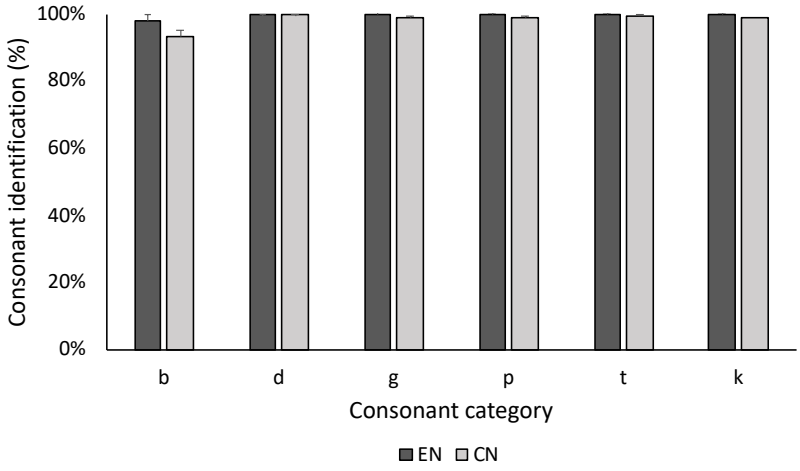

Figure 9. Identification of consonants (/b, d, g, p, t, k/) for EN and CN listeners in quiet. Error bars indicate standard error. *$p < 0.05$.

4.2.1.2 Identification of unmodified consonants in noise

The identification percentages of unmodified consonants in noise were shown in Figure 10. A four-way ANOVA (within-subjects factors: SNR, noise type and consonant category; between-subjects factor: listener group) was conducted. As a result, the main effects of SNR ($F_{1, 34} = 189.79$, $p < 0.01$, $\eta_p^2 = 0.85$), noise type ($F_{1, 34} = 82.44$, $p < 0.01$, $\eta_p^2 = 0.71$), consonant category ($F_{5, 170} = 44.46$, $p < 0.01$, $\eta_p^2 = 0.57$) were significant. Like vowel identification, lower consonant identification was found at the SNR of -15 dB than -10 dB, and in 6-TB than in LTSSN (both *ps* < 0.01). The main effect of listener

group was not significant ($F_{1, 34} = 1.16$, $p > 0.05$, $\eta_p^2 = 0.29$), suggesting that EN and CN listeners had comparable performance in overall consonant identification in noise.

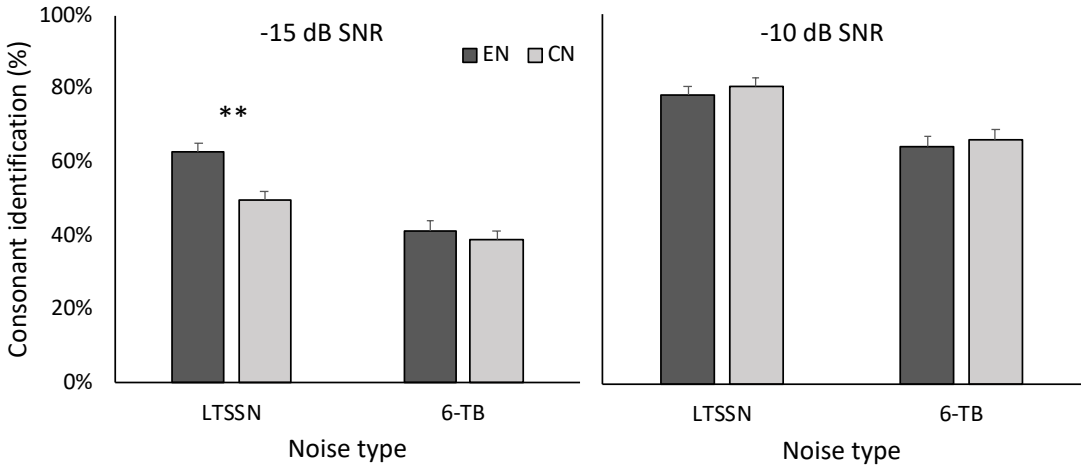

Figure 10. Identification of unmodified consonants in noise in LTSSN and 6-TB at the SNRs of -15 dB (left) and -10 dB (right). Error bars indicate standard error. **$p < 0.01$.

The significant interaction effects for consonant identification in noise were listed in Table 10. The simple effect analysis for the interaction of SNR × noise type × listener group showed EN listeners had better performance in consonant identification than CN listeners in LTSSN at the SNR of -15 dB ($p < 0.01$), but the two groups performed comparably in other noise conditions. In addition, the significant interaction effect of noise type × consonant category × listener group ($F_{5, 170} = 3.97$, $p < 0.01$, $\eta_p^2 = 0.1$)

suggested EN listeners had the better identification of /d/ in LTSSN and /k/ in 6-TB than CN listeners (both *ps* < 0.01).

Table 10: Significant interaction effects on consonant identification in noise

	Interaction effect	*F*	*p*	η_p^2
Two-factor	SNR × listener group	8.04	0.008	0.19
	SNR × consonant category	11.89	0.000	0.26
	Noise type × consonant category	30.02	0.000	0.47
Three-factor	SNR × noise type × listener group	6.75	0.014	0.17
	Noise type × consonant category × listener group	3.88	0.002	0.10
Four-factor	SNR × noise type × consonant category	8.66	0.000	0.20

4.2.2 The effect of F2 enhancement on consonant identification in quiet and noise

4.2.2.1 The effect of F2 enhancement on consonant identification in quiet

Figure 11 showed the percentage correctness of consonant identification with and without F2 enhancement in quiet. A three-way (within-subjects factors: enhancement and consonant category; between-subjects factor: listener group) ANOVA was conducted to explore the effect of F2 enhancement in the quiet condition. The main effect of

enhancement was not significant ($F_{1, 35} = 1.22$, $p > 0.05$, $\eta_p^2 = 0.04$), i.e., F2 enhancement did not improve consonant identification in quiet ($p > 0.05$). In addition, there were no significant interaction effects of listener group × enhancement, consonant category × enhancement, and listener group × consonant category × enhancement (all *ps* > 0.05).

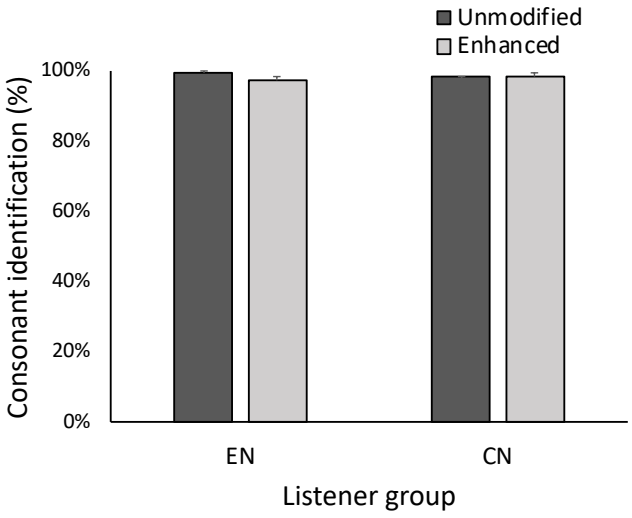

Figure 11. Identification of unmodified and enhanced consonants in quiet for EN and CN listeners. Error bars indicate standard error.

4.2.2.2 The effect of F2 enhancement on consonant identification in noise

To test the effect of F2 enhancement on consonant identification in noise conditions, a five-way (within-subjects factors: enhancement, SNR, noise type and vowel category; between-subjects factor: listener group) ANOVA was conducted. All main effects were significant (all *ps* < 0.01), including the factor of enhancement ($F_{1, 34} =$

23.33, $p < 0.01$, $\eta_p^2 = 0.41$). In general, the identification of stop consonants with F2 enhancement was better than those without enhancement ($p < 0.01$).

The significant interaction effects of F2 enhancement and other factors were shown in Table 11. The simple effect analysis for the interaction effect of enhancement × consonant category suggested that F2 enhancement significantly improved the identification of /b, d, p/ (all *ps* < 0.01), whereas F2 enhancement also considerably reduced the identification of /k/ ($p < 0.01$). In addition, the interaction effect of enhancement × SNR × noise type × listener group was significant. As was shown in Figure 12, EN listeners gained significant improvements from F2 enhancement in LTSSN and 6-TB at the SNR of -15 dB and LTSSN at -10 dB SNR (*ps* < 0.05). On the other hand, as shown in Figure 13, F2 enhancement only improved consonant identification for CN listeners in LTSSN at the SNR of -15 dB ($p < 0.01$). In addition, although the interaction effect of enhancement × noise type × consonant category × listener group was significant, EN and CN listeners showed improvements with the same conditions of noise type and consonant category. Both EN and CN listeners gained substantial improvements in identification of /b, d, p/ in LTSSN and /p/ in 6-TB (all *ps* < 0.05) and suffered an identification decline of /k/ in both LTSSN and 6-TB (all *ps* < 0.05). However, the improvement/decline amount of benefit on these consonants might differ between the two groups.

Table 11: Significant interaction effects with the factor of enhancement on consonant identification in noise

	Interaction effect	F	p	η_p^2
Two-factor	Enhancement × consonant category	48.34	0.000	0.59
Three-factor	Enhancement × SNR × consonant category	8.81	0.000	0.21
	Enhancement × noise type × consonant category	19.44	0.000	0.36
Four-factor	Enhancement × SNR × noise type × listener group	5.95	0.022	0.15
	Enhancement × noise type × consonant category × listener group	4.57	0.000	0.12
	Enhancement × SNR × noise type × consonant category	7.45	0.000	0.18

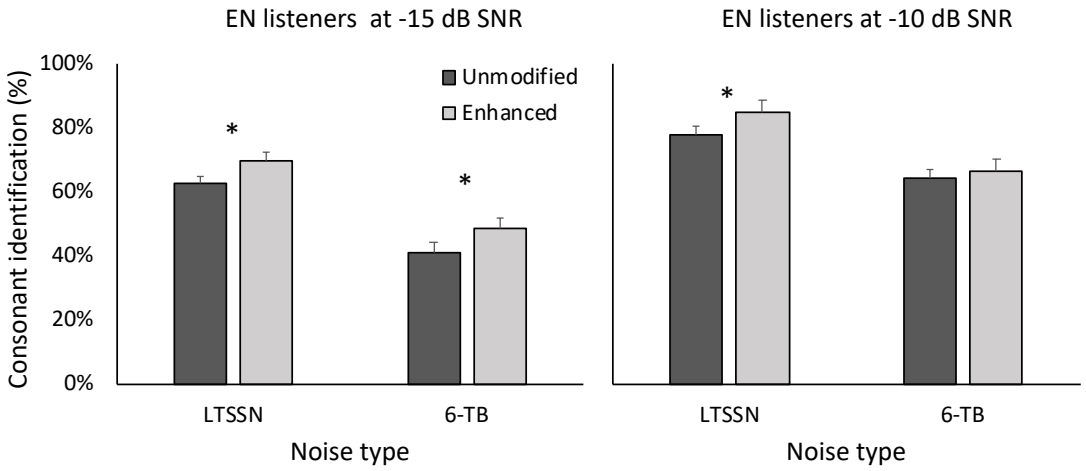

Figure 12. Identification of unmodified and enhanced consonants in LTSSN and babble noise at the SNRs of -15 dB (left) and -10 dB (right) for EN listeners. Error bars indicate standard error. *$p < 0.05$.

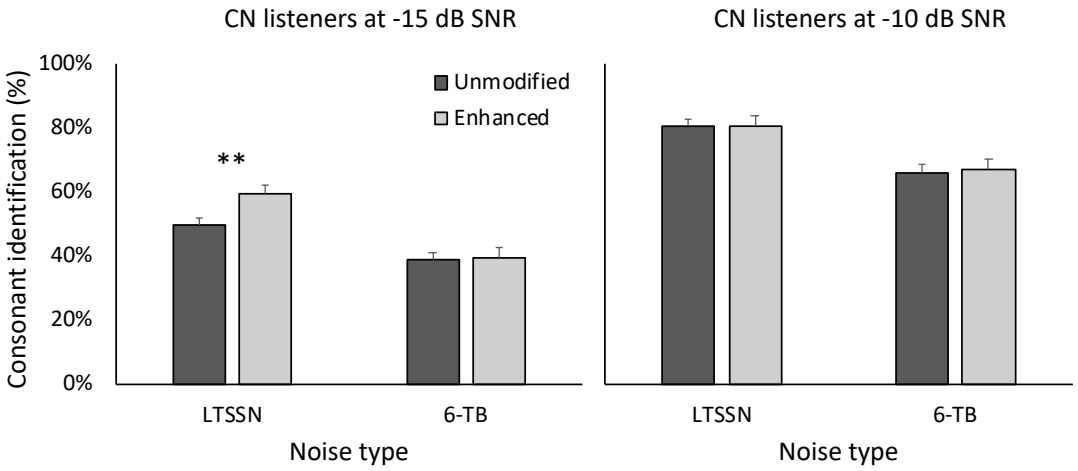

Figure 13. Identification of unmodified and enhanced consonants in LTSSN and babble noise at the SNRs of -15 dB (left) and -10 dB (right) for CN listeners. Error bars indicate standard error. $^{**}p < 0.01$.

The effects of F2 enhancement varied on the identification of different consonants (e.g., improvements for /b, d, p/ and reduction for /k/), and a three-way ANOVA (within-subjects factors: enhancement, articulation place, and voice) was conducted to explore whether the effects of F2 enhancement depend on the articulation place, voice or both. There were three levels in the factor of articulation place, including bilabials (i.e., the average percentage of /p, b/), alveolars (i.e., the average percentage of /t, d/), velars (i.e., the average percentage of /k, g/). The factor of voicing contained voiced (i.e., average percentage of /b, d, g/) and voiceless levels (i.e., average percentage of /p, t, k/). The main effects of F2 enhancement ($F_{1, 35} = 20.82$, $p < 0.01$, $\eta_p^2 = 0.37$), articulation place

($F_{2, 70} = 32.91$, $p < 0.01$, $\eta_p^2 = 0.55$) and voicing ($F_{1, 35} = 32.91$, $p < 0.01$, $\eta_p^2 = 0.49$) were significant. The *post hoc* comparisons for articulation place suggested the identification of alveolar consonants were significantly higher than bilabial and velar consonants (both *ps* < 0.01). At the same time, there was no significant difference between the latter two ($p > 0.05$). In addition, the significant main effect of voicing showed better identification of voiceless consonants than voiced consonants ($p < 0.01$). The interaction effects of enhancement × articulation place was significant ($F_{2, 70} = 65.16$, $p < 0.01$, $\eta_p^2 = 0.65$), while the interaction effect of enhancement × voicing was not ($F_{1, 35} = 0.24$, $p > 0.05$, $\eta_p^2 = 0.01$). Figure 14 illustrated the effects of F2 enhancement on bilabials, alveolars and velars. The simple effect analysis of enhancement × articulation place interaction showed significant improvements with F2 enhancement for bilabials and alveolars (both *ps* < 0.01), but a decline for velars ($p < 0.01$).

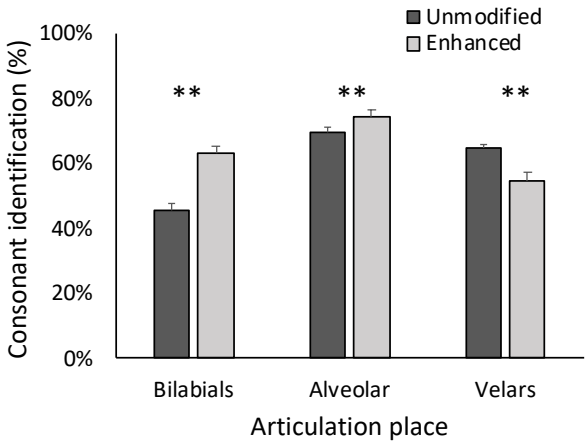

Figure 14. Identification of unmodified and enhanced bilabials, alveolars, velars in noise. Error bars indicate standard error. $^{**}p < 0.01$.

4.2.2.3 Effect of F2 enhancement on consonant confusion matrix in noise

Table 12 showed the effect of F2 enhancement on consonant confusion matrix (i.e., differences of confusion matrices in noise with and without F2 enhancement) for all the listeners. Both EN and CN listeners showed similar effects of F2 enhancement across consonant categories, e.g., identification of /b, d, p/ were significantly improved while those of /k/ were reduced. The analysis of confusion matrix suggested that increased identification of /b, p/ with F2 enhancement mainly came from the reduced confusions with the consonants at different articulation places rather than the voiced-voiceless confusions. Differently from /b, p/, the improved identification of /d/ (i.e., increase of 7.1%) was derived from the reduced confusion with its voiceless counterpart /t/ (i.e.,

reduction of 6.4%). In addition, the identification of /k/ was reduced with F2 enhancement (i.e., reduction of 18.1%) primarily due to the increased confusions with the corresponding voiced stop /g/ (i.e., increase of 17.5%).

Table 12: Effect of F2 enhancement on consonant confusion matrix in noise

Response / Target	/b/	/d/	/g/	/p/	/t/	/k/
/b/	**6.7%**	-3.9%	-7.6%	7.7%	-1.3%	-1.5%
/d/	-0.9%	**7.1%**	1.4%	-0.5%	-6.4%	-0.6%
/g/	-0.7%	-1.1%	**-1.1%**	1.7%	1.6%	-0.6%
/p/	2.1%	-2.1%	-5.5%	**27.2%**	-7.7%	-14.0%
/t/	-0.9%	0.2%	-0.6%	-0.4%	**2.3%**	-0.5%
/k/	0.4%	-0.1%	17.5%	0.5%	-0.2%	**-18.1%**

4.3 The amount of benefit from F2 enhancement

The amount of perceptual benefit was computed by subtracting the identification score of unmodified speech without F2 enhancement from that of enhanced speech in the same listening condition. For the purpose of comparing the enhancement benefit between vowels and consonants, the amount of benefit on vowel and consonant identification was computed as the average over the five vowels and six consonants, respectively. As the significant improvement occurred at various conditions of the factors of phonetic type,

listening conditions and language experience, a four-way ANOVA (within-subjects factors: phonetic type, SNR, noise type; between-subjects factor: listener group) was conducted to investigate whether the amount of benefit would be affected on the three factors, as well as their interactions. As a result, the main effect of SNR was significant ($F_{1, 34} = 19.72$, $p < 0.01$, $\eta_p^2 = 0.37$), while the other main effects were not (all $ps > 0.05$). In particular, the perceptual benefit from F2 enhancement was significantly higher at the SNR of -15 dB than at the SNR of -10 dB ($p < 0.01$).

Table 13 displayed the significant interaction effects of the amount of benefit. As shown in Figure 15, the simple effect analysis on phonetic type × listener group suggested that the amount of help from F2 enhancement was higher for the CN listeners than the EN listeners in vowel identification ($p < 0.01$). At the same time, it was similar between the two groups in consonant identification ($p > 0.05$). In addition, the analysis compared by phonetic type suggested more benefits on vowel identification than on consonant identification for the CN listeners ($p < 0.01$), but comparable benefits between vowel and consonant identification for EN listeners ($p > 0.05$).

Table 13: Significant interaction effects of the amount of benefit from F2 enhancement

	Interaction effect	F	p	η_p^2
Two-factor	Phonetic type × listener group	7.82	0.008	0.19
	Phonetic type × SNR	6.23	0.018	0.16
	Phonetic type × noise type	11.61	0.002	0.25
	SNR × noise type	10.49	0.003	0.24
Three-factor	Phonetic type × SNR × noise type	8.01	0.000	0.19

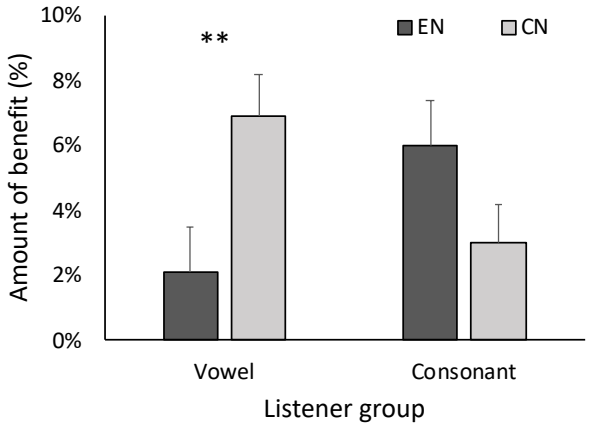

Figure 15. The amount of perceptual benefit for EN and CN listeners in vowel and consonant identification. Error bars indicate standard error. $^{**}p < 0.01$.

The interaction effect of phonetic type × SNR × noise type was significant ($F_{1, 34}$ = 8.01, $p < 0.01$, $\eta_p^2 = 0.19$). The simple analysis compared by noise type found a higher benefit in 6-TB than in LTSSN at the SNR of -10 dB in vowel identification (shown in Figure 17; $p < 0.01$). Meanwhile, the amount of benefit was comparably between LTSSN and 6-TB in all other conditions (e.g., vowel identification at -15 dB SNR and consonant identification at both -10 dB and -15 dB SNR) (all *ps* > 0.05).

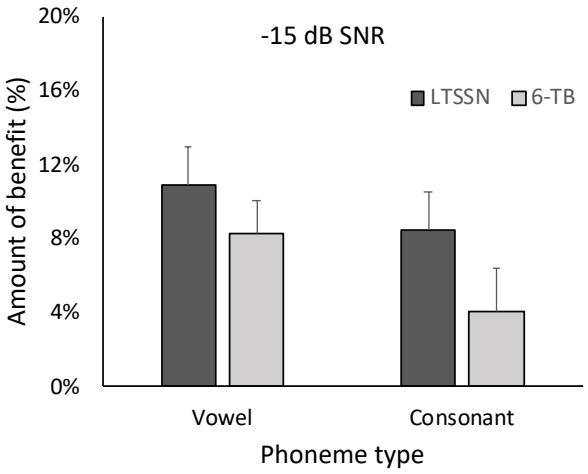

Figure 16. The amount of perceptual benefit in LTSSN and 6-TB at the SNR of -15 dB in vowel and consonant identification. Error bars indicate standard error.

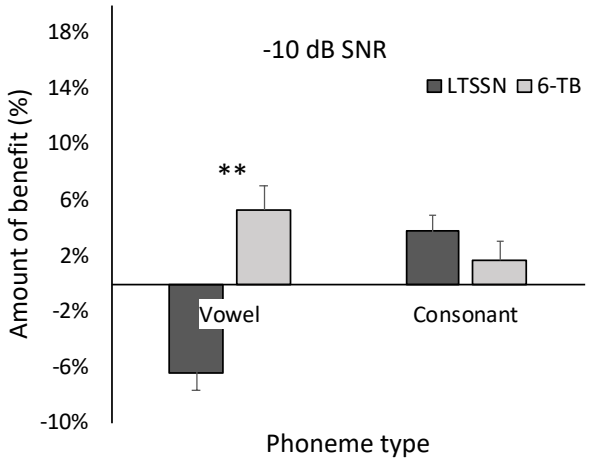

Figure 17. The amount of perceptual benefit in LTSSN and 6-TB at the SNR of -10 dB in vowel and consonant identification. Error bars indicate standard error. $^{**}p < 0.01$.

The Spearman's rank correlation analyses were conducted between self-evaluations of English proficiency (e.g., overall proficiency and the abilities of reading, speaking, listening, and writing) and the amounts of benefit (e.g., in vowel and consonant identification in quiet and noise) for CN listeners (see Table 14). The results suggested no significant correlation between the two factors across quiet and noise listening conditions.

Table 14: Spearman's correlation coefficients (r) between the amount of benefit with self-evaluations of English proficiency in vowel and consonant identification

	Vowel identification		Consonant identification	
	Quiet	Noise	Quiet	Noise
Overall proficiency	0.33	0.08	0.15	-0.16
Reading	-0.01	0.07	0.15	-0.05
Speaking	0.37	0.02	0.19	-0.36
Listening	0.35	0.21	0.07	-0.41
Writing	0.13	-0.38	0.10	0.06

Overall, the results suggested the amount of benefit from F2 enhancement was generally comparable between vowel and consonant identification, between LTSSN and 6-TB, and between native and nonnative listeners. A higher amount of benefit was found at -15 dB SNR than at -10 dB SNR, as well as for nonnative listeners in vowel identification than for native listeners.

Chapter 5: Discussion

The current study investigated whether F2 enhancement could benefit English phonetic perception for native and non-native listeners in various listening conditions, particularly whether and how the benefits of F2 enhancement were dependent on phonetic type, listening conditions, and listeners' language background. In this section, F2 enhancement on vowel and consonant identification in noise is described, respectively, and followed by a general discussion.

5.1 Vowel identification

5.1.1 Identification of unmodified vowels in quiet and noise

This study investigated the effects of noise masking on unmodified vowel identification as a base line. The average vowel identification across vowel categories was 99.6% for native listeners and 73.8% for nonnative listeners in quiet conditions. As shown in Figure 5, vowel identification for both native and nonnative listeners obviously dropped in noise conditions as expected, listeners in the two groups suffered more noise-masking at the SNR of -15 dB than the -10 dB SNR. In addition, noise type is another critical factor that influences vowel identification. Two types of noise, LTSSN, and 6-TB, served as the masker in the current study. 6-TB is a speech noise simulating six persons were talking simultaneously, while LTSSN is a stationary nonspeech noise with the spectrally matched with 6-TB. With the temporal dips and speech contents, 6-TB may

provide the release of energetic masking from the temporal fluctuation of babble and informational masking, both of which are missing in LTSSN. As a result, vowel identification for both native and nonnative listeners was comparable in LTSSN and 6-TB at the SNR of -15 dB, while scores were lower in 6-TB than in LTSSN at the -10 dB SNR. These results might because at the SNR of -15 dB, the amount of informational masking in 6-TB was approximately equivalent to the release of energetic masking from temporal dips in babble. In contrast, at the SNR of -10 dB, the amount of informational masking might be more significant than the energetic masking release in speech noise. As the SNRs are at the easy to moderately challenging levels (e.g., 10 dB to -10 dB), it becomes more difficult for listeners to separate speech signals from babble noise, e.g., listeners' confusion on what speech signal was and what background babble was.

This study also showed higher vowel identification for a native listener in both quiet and noise conditions than nonnative listeners. The native-nonnative difference of vowel identification was 25.8% in quiet; meanwhile, these differences were not obviously enlarged in noise conditions (e.g., 21.9% in LTSSN and 20.4% in 6-TB at the SNR of -10 dB and 27.4% in LTSSN and 25.2% in in 6-TB at the -15 dB SNR). Thus, nonnative listeners did not suffer extra difficulties in noise conditions than native listeners, which was somewhat consistent with the conclusion in previous studies of vowel identification in noise (M. Cooke et al., 2008; Mi et al., 2013). In addition, the nonnative disadvantages were found in the identification of all the front vowels (e.g., /æ, e, ɛ, i, ɪ/) in quiet, but only of /æ, e, ɛ, ɪ/ in noise at both -10 dB and -15 dB. The

nonnative disadvantage in identifying of /i/ (e.g., 14.3%) was smaller than other vowels in quiet conditions. It was further reduced to a nonsignificant level in background noise, which might be attributed to two reasons: first, there was a counterpart /i/ in Mandarin Chinese; second there was a relatively high identification score of /i/ in quiet for nonnative listeners (e.g., 85.1% in quiet).

5.1.2 The effect of F2 enhancement on vowel identification in quiet and noise

This study investigated the perceptual effects of F2 enhancement with the comparisons of phonetic identification scores between unmodified and enhanced speech. In general, perceptual improvement from F2 enhancement was found in noise conditions but not in quiet, consistent with the findings in our previous studies with formant frequency discrimination (Woodall & Liu, 2013; Guan & Liu, 2019b). Woodall and Liu (2013) suggested the F2 enhancement could not improve vowel formant sensitivity in quiet for young listeners with normal hearing, while Guan and Liu (2019b) found significant improvements in vowel formant discrimination in LTSSN. These results indicate that in quiet, the resolution of formant peaks in natural speech is good enough for young normal hearing listeners' vowel perception. Therefore, there is no need to increase spectral contrasts for formant peaks. However, noise background may significantly smear spectral resolution of vowel formants. Thus, spectral enhancement may be beneficial for vowel perception in noise.

In this study, F2 enhancement significantly improved vowel identification in most of the noise conditions. Furthermore, vowel identification became significantly better with F2 enhancement in both types of noise except the LTSSN at the SNR of -10 dB in which vowel identification was reduced considerably. These results suggested that F2 enhancement could lead to either positive or negative effects. On the one hand, the enhancement algorithm strengthened the audibility of F2 with higher spectral contrasts, which was potentially beneficial for vowel processing as a positive effect. However, on the other hand, the enhancement also changed the amplitude ratios of formants (e.g., F1/F2) and spectral tilt, which might lead to speech distortion and reduced intelligibility (Liu & Eddins, 2008).

The overall effect of F2 enhancement on vowel perception may depend on the type of noise and SNR, as suggested in this study. For very challenging SNRs such as -15 dB, the audibility problem of formant peaks is usually dominant for vowel identification, which can be significantly compensated by F2 enhancement. In contrast, speech distortion and increased spectral tilt of vowels by F2 enhancement may be ignored due to high-level noise. On the other hand, at medium-level noise, the benefits brought by F2 enhancement may be reduced and even overturned because of speech quality distortion that was associated with F2 enhancement. In addition, in quiet conditions, speech cues are fully available to be processed. Neither enhanced spectral contrasts of F2 nor speech distortion by F2 enhancement seemed to affect speech perception for native and non-native listeners.

Both native and nonnative listeners benefited from F2 enhancement in vowel identification in most noise conditions (e.g., in LTSSN and 6-TB at -15 dB SNR, and in 6-TB at -10 dB SNR). Although CN listeners had lower sensitivity to formant frequency change than EN listeners (Liu et al., 2012), enhanced F2 peaks improved their English vowel identification in noise, which primarily relies on the processing of vowel formants. Combined with the findings of perception training (Hu et al., 2016; Ylinen et al., 2010) and second language learning experience (Flege et al., 1997; Hsieh & Pan, 2010) on formant processing for nonnative listeners, it was suggested that their dependence on duration in vowel identification in noise should be attributed to their difficulties in spectral processing. Therefore, either L2 speech training/experience or F2 enhancement of speech materials could improve the effectiveness of using formant cues for nonnative listeners.

In addition, the effects of F2 enhancement varied across vowel categories. Native listeners gained improvements from F2 enhancement in identifying /e, i/ in noise, and the decline in /æ, ɛ, ɪ/ identification. An explanation was proposed as the acoustic features of tense and lax vowels. Vowels can be categorized into tense or lax vowels based on the tension degree in the tongue muscles, especially those for the bunching up of the tongue lengthways (Durand, 2005). Lax vowels tend to be centralized in the pronunciation space compared with tense vowels. Tense vowels tend to have longer durations, less formant frequency changes and dynamic formant frequency movements in the vowel duration. Meanwhile, the duration is generally shorter, and formant frequency change and dynamic

formant frequency movement are greater for tense vowels (Leung, Jongman, Wang, & Sereno, 2016). In the current study, the identification of tense vowels, /e, i/, in noise were improved with F2 enhancement, whereas the identification of lax vowels, /æ, ɛ, ɪ/, were not. One possibility is that more stable formant information and longer duration of tense vowels may boost the perceptual effects of F2 enhancement. On the other hand, the temporally more fluctuating formant trajectory and shorter duration of lax vowels may reduce the perceptual benefits of F2 enhancement. Like native listeners, nonnative listeners benefited from tense vowels but suffered decreased identification for lax vowels by F2 enhancement. However, it should also be noted that reduced identification for nonnative only occurred for /ɪ/ instead of /æ, ɛ/. The results suggested nonnative listeners suffered less from negative effects of F2 enhancement on some lax vowels (e.g., /æ, ɛ/), different from native listeners. One possible reason is that the priority of nonnative listeners in L2 phonetic perception was to identify the target speech signal, which is more important than the quality of speech. That is, nonnative listeners may tolerate the distortion of speech, if any, caused by F2 enhancement as long as they can recognize the target sound. In addition, the low identification score of /ɪ/ (e.g., 30.2%), compared to that of /æ/ (e.g., 71%) and /ɛ/ (e.g., 49.4%), might significantly limit the benefits of F2 enhancement, i.e., the ceiling identification score of /i/ identification was quite low.

5.2 Consonant identification

5.2.1 Identification of unmodified consonants in quiet and noise

The stop consonant identification was at high levels for native and nonnative listeners in quiet, e.g., 99.6% for EN listener and 98.1% for CN listeners. As shown in Figure 10, the identification of stop consonants for both native and nonnative listeners was also greatly affected by adverse background noise. In addition, the effects of noise masking on consonant identification seemed even higher or at least like vowel identification in noise conditions. The results were inconsistent with previous studies that suggested less susceptibility of noise masking for consonant identification than for vowel identification (Mi et al., 2013; Parikh & Loizou, 2005; Tao et al., 2018). The previous inconsistency might be partly attributed to the vowel categories as the stimuli, e.g., the current study selected front vowels as stimuli in vowel identification. In contrast, more types of vowels (e.g., front, central and back vowels) were served in previous studies. Liu and Jin (2019) found greater slopes of back and central vowels than front vowels in psychometric functions of vowel identification, indicating that front vowels suffer less noise masking with the decrease of SNR than central and back vowels. Another possibility might be the different noise conditions, e.g., -5 dB SNR in a study by Parikh & Loizou (2005) and 12-talker babble used in the study by Mi et al. (2013) and Tao et al. (2018), which might have different masking effects on phonetic identification compared with the noise conditions of LTSSN and 6-TB at -10 and -15 dB in the current study. In addition, the total amount of noise masking on consonant identification in 6-TB was higher than LTSSN across the two SNRs, suggesting that the negative effect of informational masking in 6-TB might be greater than the benefits of temporal dip

listening in babble. Although those nonnative listeners showed lower consonant identification than native listeners in quiet conditions, the performances of both native (99.6%) and nonnative (98.1%) listeners were at quite high levels. Moreover, native and nonnative listeners showed comparable consonant identification generally in noise conditions. The results suggested that CN listeners have native-like performance in English stop consonant identification, possibly because all these stops have the counterparts in Mandarin Chinese.

5.2.2 The effect of F2 enhancement on consonant identification in quiet and noise

F2 enhancement generally improved consonant identification in noise, while the improvement was different between native and nonnative listeners in various listening conditions. Native listeners gained significant improvements in most noise conditions (e.g., LTSSN and 6-TB at -15 dB SNR, as well as LTSSN at -10 dB). In contrast, the significant benefits for nonnative listeners only occurred in nonspeech noise at low SNRs (e.g., LTSSN at -15 dB SNR). Such differences between native and nonnative listeners in the number of noise conditions where the perceptual benefit of F2 enhancement was found were possibly due to the difference in perceptual weights of acoustic cues between native and nonnative listeners. Formant transitions and transient release burst are two critical cues in the identification of stop consonants, and their perceptual weights are reciprocal, e.g., the increase in the perceptual weight of one cue is associated with the decrease of the weight of the other cue (Dorman et al., 1977; Story & Bunton, 2010). One

possibility is that for stop consonant perception in noise, EN listeners rely heavily on formant transitions for stop consonant perception in noise, while CN listeners put less perceptual weights on formant transition.

The effects of F2 enhancement also varied across consonant categories. F2 enhancement improved the identification of /b, d, p/, but reduced the identification of /k/ in noise conditions. As shown in Table 15, the six English stops are divided into bilabials (e.g., /p, b/), alveolars (e.g., /t, d/), and velars (e.g., /k, g/) based on the articulation place, as well as into voiced (e.g., /b, d, g/) and voiceless consonants (e.g., /p, t, k/) according to whether the vocal cords vibrate in the pronunciation (Ladefoged & Johnson, 2014; Olive, Greenwood, & Coleman, 1993). Further analysis suggested that the perceptual effects among consonant categories depend on the articulation place instead of voicing. F2 enhancement could improve the identification of bilabial and alveolar sounds but not for velar identification. The spectral energy distribution of F2 transition and the transient release burst of consonants may account for the difference in the F2 enhancement effect across consonant categories. The energy of bilabial consonants is predominantly distributed at low frequencies (500 - 800 Hz and up to 1500 Hz) and alveolar stops usually have more energy at high frequencies (4000 Hz and above) (Blumstein & Stevens, 1979; Halle, Hughes, & Radley, 1957). The energy concentration areas of bilabial and alveolar in the spectrum are distant from the F2 region of the surrounding vowel /a/ (e.g., about 1200 Hz). On the other hand, the velar stops usually have a compact spectrum in the mid-frequency range from 1,000 - 3, 000 Hz (Blumstein &

Stevens, 1979), spectrally overlapped with the F2 region of surrounding vowel /a/. The enhanced speech with enhanced F2 peaks might lead to greater forward and/or backward masking to the release burst located in the middle of the CVC stimulus than unmodified speech. Thus, the forward and backward masking from the enhanced F2 peaks is expected to be more significant for velar stops due to the spectral overlap. Furthermore, the analysis on consonant confusion matrix changes suggested that the F2 enhancement would improve the distinction of articulation places for bilabials. At the same time, it was primarily beneficial for the voiceless-voiced distinguishment for the alveolars (e.g., /d/). In addition, the identification of velar /k/ was reduced with F2 enhancement mainly due to more voiceless-voiced confusions (e.g., /k/ versus /g/). An assumption was proposed that the weights of F2 transitions and other acoustic cues (e.g., transient burst release) might be different among bilabials, alveolars and velars, e.g., F2 transitions might play a more critical role in articulation place distinguishing for bilabials, while in voiceless-voiced distinction for alveolars and velars, however, more studies are still in need to examine the speculations.

Table 15: The classification of English stop consonant

Voicing	Place of articulation						
	Bilabial	Labiodental	Dental	Alveolar	Palatal	Velar	Glottal
Voiceless	p			t		k	
Voiced	b			d		g	

5.3 The amount of benefit with the factors of phonetic type, noise conditions and language experience

The significant benefits of F2 enhancement in noise were presented in various experimental conditions of phonetic type, noise conditions (e.g., SNR and noise type) and listeners' language experience. The current study further explored whether the amount of benefit varied across these factors and their interaction effects. Overall, the perceptual benefits were comparable between vowels and consonants, between LTSSN and 6-TB, and between native and non-native listeners. In contrast, the benefits varied between two SNRs and across some interactions of the three factors.

The benefits of F2 enhancement were comparable between vowel and consonant identification in most noise conditions (e.g., babble and nonspeech noise at the SNR of -15 dB and babble at the SNR of -10 dB). The results were inconsistent with the hypothesis that more benefits would occur in vowel identification with a higher perceptual weight of F2 information. It might be partly attributed to the offsetting of the

positive and negative effects of F2 enhancement. As discussed in 5.1.2 and 5.2.2, F2 enhancement might have both positive and negative effects on speech perception in noise. The perceptual outcome might primarily depend on the combination of both positive and negative effects. For vowel identification, F2 enhancement may bring both positive (e.g., better local SNRs for F2) and negative (e.g., shallower spectral tilt) effects. On the other hand, for consonant identification. F2 enhancement improves the audibility of F2 transition, possibly forcing listeners to use the F2 transition cue to perceive stop consonants, particularly when release bursts are primarily masked by noise.

The amount of benefit from F2 enhancement depended on SNR, and more benefits occurred at the SNR of -15 dB. Thus, it suggested the F2 enhancement is more suitable in very challenging listening conditions, consistent with the study by Chen et al. (2012, 2013), who reported that dynamic spectral enhancement could improve speech perception at -6 dB SNR but not at -3 dB SNR.

The amount of benefit from F2 enhancement was similar between LTSSN and babble in general, e.g., for vowel identification at -15 dB SNR and consonant identification at both -15 dB and -10 dB SNRs. LTSSN has energetic masking on speech perception, while 6-TB at the same SNRs usually has less energetic masking but with more informational masking. The comparable amount of benefit between the two types of noise suggested that F2 enhancement could undoubtedly reduce the energetic masking of noise and may lower the informational masking of babbles on phonetic identification.

However, further studies are needed to investigate how F2 enhancement improves listeners' capacity against noise's energetic and informational masking.

Generally, the enhanced benefits were comparable or different for native and nonnative listeners, depending on the phonetic type. Compared to native listeners, nonnative listeners gained more benefits in vowel identification, but similar benefits in consonant identification, possibly due to nonnative disadvantages and perceptual weight of F2 information on vowel and consonant identification. Previous studies suggested nonnative listeners had considerable disadvantages in formant processing Cutler et al., 2005; Flege et al., 1997; Kondaurova & Francis, 2008; Liu et al., 2012; Mi et al., 2016; Morrison, 2009; Tyler & Cutler, 2009; Wang, 2006; Ylinen et al., 2010), as well as the capacities against energetic and informational masking (Guan, Liu, Tao, Li et al., 2015; Guan, Liu, Tao, Mi et al., 2015; Mi et al., 2013; Stuart et al., 2010; Van Engen, 2010). The enhancement algorithm of this study strengthened the spectral cue of F2 that was degraded in noise. Nonnative listeners showed substantial disadvantages in vowel identification in noise but a much smaller disadvantage in consonant identification. At the phonetic level, non-native listeners' challenge is vowel perception rather than consonant perception, leaving much greater room for vowel identification to be improved than for consonant identification. In addition, the amount of benefit in phonetic identification was not significantly related to the self-evaluated English proficiency for nonnative listeners, suggesting that there were some factors other than the phonetic processing to account for the individual variability in F2 benefits for nonnative listeners.

5.4 General discussion

The current study investigated the interaction effects of F2 enhancement phonetic type, listening conditions (e.g., factors of SNR and noise type) and listeners' language experience. Overall, perceptual improvements by F2 enhancement were generally found across these factors, i.e., F2 enhancement increased the identification of vowels and consonants for both native and non-native listeners in speech and nonspeech noise at two challenging SNRs.

The significant improvement in 6-TB in the current study was inconsistent with the study by Guan and Liu (2019a). With the task of word recognition, they found that the F2 enhancement could improve speech perception for the older listeners with normal hearing and hearing impairment in two-talker babble, but not in 6-TB. Three possible reasons were proposed for the different findings between the current and previous studies. The first possibility was the different enhanced scales for F2 between the two studies. The F2 was improved by 9 dB in the study by Guan et al. (2019a), while the improved degree reached to 12 dB in the current study. This study's more prominent F2 enhancement in this study could lead to the improvements in challenging noise conditions such as in 6-TB with less temporal dips and more speech contents in background noise - the second one concerned with the age factor for participants. Guan and Liu (2019a) recruited the elder listeners with the average age of 60.4 years old, while young

participants were chosen in the current study with an average age of 20.7 years. The perceptual effects of F2 enhancement might be different for young and elder listeners with normal hearing. Third, the different findings might be partially attributed to the different various significant information-masking underlying word and phonetic identification mechanisms. Word recognition usually suffers more from semantic interference from background babbles (Carhart, Johnson, & Goodman, 1975), while the misallocation of speech cues (e.g., formants) might be the significant type of informational masking for phonetic perception (Simpson & Cooke, 2005). The F2 enhancement might have different effects on the mechanisms of informational masking. Further studies are needed to examine how F2 enhancement interacts with the informational masking of MTBs.

Although different speech perception tasks, listening conditions, listeners' hearing status, and language experience were used in the current and previous studies of speech enhancement, it is worth comparing the perceptual effects across these studies. On the one hand, the benefit amount from F2 enhancement appeared to be higher compared with that of previous research. On the other hand, the traditional enhancement strategies like spectral sharpening and expansion resulted in small (3-5%) or even no benefit on speech perception in noise in most of the previous studies (Bunnell, 1990; Franck et al., 1999a; Rout, 2006; Stone & Moore, 1992; Summerfield et al., 1985). In this study, F2 enhancement increased 10.9% in vowel identification and 8.4% in consonant identification in speech noise at low SNRs (e.g., -15 dB). Furthermore, speech

intelligibility improved with the dynamic enhancement reached 8%, and even 14% after selecting of the best parameters individually for each participant; however, the improvement only occurred in LTSSN and was not found in MTB (Chen et al., 2012). In this study, the perceptual effects of F2 enhancement were observed at both LTSSN and 6-TB.

Despite the overall perceptual benefit of F2 enhancement in this study, the perceptual effect of F2 enhancement should be considered as a 'double-edged sword', instead of a one-way improvement. For example, for vowel identification, F2 enhancement was beneficial in most noise conditions for vowel identification but also resulted in an identification decline in some noise conditions, such as at the LTSSN of -10 dB SNR. F2 enhancement resulted in improvements in some vowels (lax vowels) and consonants (e.g., bilabials and alveolars), but reduction in other vowels (tense vowels) and consonants (e.g., velars). In addition, more benefits from F2 enhancement were found at a low SNR (e.g., -15 dB SNR), as well as for nonnative listeners in vowel identification with disadvantages in speech perception in noise. These results possibly suggested F2 enhancement is more helpful in challenging listening conditions and for listeners with more significant perceptual difficulties.

5.5 Limitations of this study

There are still some limitations of current research to be further explored in future studies. First, instead of all categories of English vowels and consonants, the present

study selected front vowels to reduce the effect of dialect (e.g., significant /ɔ/-/ɑ/ confusion in Texans) and stop consonants that depend on formant transitions heavily for perception. Second, it is not clear whether the effects of F2 enhancement will remain similar in identifying of other vowels (e.g., central, and back vowels) and consonants (e.g., nasals, fricatives, affricates, liquids, and glides). In addition, it is unknown whether the effects of F2 enhancement are underestimated or overestimated without inter-categories confusions (e.g., the confusions between front vowels and central or back vowels).

Second, the factor of the enhanced scale was not systematically investigated in the current study, and the perceptual effects with different enhanced degrees remain unclear. In this study, the F2 enhancement might have both positive and negative effects on phonetic identification in noise. Further studies are required to investigate the best-enhanced scales (e.g., 6, 9 and 12 dB) with the maximum benefits in noise conditions.

Third, the effects of F2 enhancement on energetic and informational masking were preliminarily discussed with the comparison of LTSSN and 6-TB. However, the amounts of energetic and informational masking in speech noise could not be clearly separated. Therefore, future studies are needed to quantify the energetic and informational masking of multi-talker babbles by including babble-modulated noise that spectrally and temporally matched with babble.

5.6 Future research directions

The current and previous studies investigated the effects of F2 enhancement with different speech materials (e.g., vowel, consonant, and words), noise conditions (e.g., speech and nonspeech noise at different SNRs), populations (e.g., young native and nonnative listeners with normal hearing, elderly listeners with normal hearing and hearing loss) and experimental tasks (e.g., phonetic identification, word recognition, and formant frequency discrimination). Future studies are to focus on other populations and speech materials.

First, the current study found that nonnative listeners gained more benefits in vowel identification than native listeners, suggesting that the peripherally enhanced information might be used to compensate for their disadvantages in phonological processing at the central level. Future studies may further investigate whether the compensation from F2 enhancement will also present in the population with processing deficit (e.g., major auditory processing disorders and those with both hearing and phonological difficulties (e.g., nonnative listeners with hearing loss).

Second, the current strategy focused on enhancing F2 only, while F3 is also considered as a useful spectral cue especially for consonants (Alwan, 1992; Harris et al., 1958; Story & Bunton, 2010). In addition, the intensity of F3 is even lower than F2 in general, which is more susceptible to noise masking. Therefore, it is an interesting topic to investigate whether the enhancement of both F2 and F3 will lead to more significant benefits compared with F2 enhancement only.

Third, F2 enhancement may also be beneficial in more situations, such as perception of muffled speech. Currently, face masks are widely used to slow the spread of the COVID-19. However, face masks, especially N95 respirators and face shields, usually muffle speech with high-frequency attenuations (Corey, Jones, & Singer, 2020; Goldin, Weinstein, & Shiman, 2020), leading to more difficulties for nonnative listeners, as well as the listeners with hearing problems. Therefore, formant enhancement is expected as a potential solution by strengthening the attenuated spectral cues.

5.7 Potential application for F2 enhancement

Based on the current study's findings, listeners with normal hearing gained improvements in phonetic identification in various noise conditions. Therefore, the algorithm of F2 enhancement is promising to be applied in hearing devices used in adverse noise conditions, e.g., broadcast in vehicles.

Moreover, combined with the previous studies, formant enhancement also has potential amplification applications for the listeners with auditory processing deficits. Formant enhancement can improve the audibility of speech in adverse listening conditions without increasing the overall intensity. Thus, it is suitable for listeners who have difficulties in speech perception in noise but cannot tolerate the excessive intensity of speech (e.g., listeners with loudness recruitment). In addition, the philosophy to reduce spectral smearing of critical formants might be an inspiration for hearing device manufacturers and audiologists.

Chapter 6: Conclusion

In conclusion, the current study found that F2 enhancement by 12 dB could improve speech perception in challenging noise in general and in various conditions of phonetic type, noise conditions and language experience. Furthermore, the results indicated the F2 enhancement could improve vowel and consonant identification for native and nonnative listeners in various listening conditions of speech and nonspeech noise at -10 dB and -15 dB SNR. In addition, the benefit amount from the F2 enhancement depends more on SNR, as well as the interaction of phonetic type and language experience, e.g., more benefits were found at the very challenging SNR (i.e., -15 dB), as well as for nonnative listeners in vowel identification.

Appendix 1. Questionnaire for Bilingual Speakers

This questionnaire is related to the amount of English you have been exposed in your life. Please choose the best answer that describes your language background.

1. Name: Age: Gender: Male / Female
2. English proficiency: CET4 Score:
3. Email address: _____
4. At what age did you first begin to learn English? And in what format (home or school)?

5. Rate your current overall language ability in ENGLISH

   1 = understand but cannot speak

   2 = understand and can speak with great difficulty

   3 = understand and speak but with some difficulty

   4 = understand and speak comfortably, with little difficulty

   5 = understand and speak fluently like a native speaker

6. On a scale from 1 to 5, rate your abilities in English

   (1 =poor; 2= needs work; 3=good; 4= excellent; 5= native speaker level)

   Reading = Speaking= Listening= Writing=

7. Do you have normal hearing? (Y/N)

8. Please indicate the ratio of using English and your native language (e.g., 40-60):

9. Do you have any history of speech and/or language disorders? (Y/N)

Reference

Alwan, A. (1992). The Role of F3 and F4 in Identifying Place of Articulation for Stop Consonants. Paper presented at the *Second International Conference on Spoken Language Processing*.

ANSI. (2010). American National Standard Specification for Audiometers (ANSI S3. 6-2010). In: Author New York.

Arbogast, T. L., Mason, C. R., & Kidd Jr, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America, 112*(5), 2086-2098.

Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In *Speech processing in the auditory system* (pp. 231-308): Springer.

Assmann, P. F. (1995). The role of formant transitions in the perception of concurrent vowels. *J Acoust Soc Am, 97*(1)

Baer, T., Moore, B. C., & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on

intelligibility, quality, and response times. *Journal of rehabilitation research and development, 30*, 49-49.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America, 66*(4), 1001-1017.

Boers, P. (1980). Formant enhancement of speech for listeners with sensorineural hearing loss. *IPO annual progress report, 15*, 21-28.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. *Speech perception and linguistic experience: Issues in cross-language research*, 279-304.

Brimijoin, W. O., Whitmer, W. M., McShefferty, D., & Akeroyd, M. A. (2014). The effect of hearing aid microphone mode on performance in an auditory orienting task. *Ear and hearing, 35*(5), e204.

Bruce, I. C. (2004). Physiological assessment of contrast-enhancing frequency shaping and multiband compression in hearing aids. *Physiological measurement, 25*(4), 945.

Brungart, D. S., & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *The Journal of the Acoustical Society of America, 112*(6), 2985-2995.

Bunnell, H. T. (1990). On enhancement of spectral contrast in speech for hearing-impaired listeners. *The Journal of the Acoustical Society of America, 88*(6), 2546-2556.

Carhart, R., Johnson, C., & Goodman, J. (1975). Perceptual masking of spondees by combinations of talkers. *The Journal of the Acoustical Society of America, 58*(S1), S35-S35.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34*(3), 372-387.

Chen, H. C., & Wang, M. J. (2011). An acoustic analysis of Chinese and English vowels.

Chen, J., Baer, T., & Moore, B. C. (2012). Effect of enhancement of spectral changes on speech intelligibility and clarity preferences for the hearing impaired. *J Acoust Soc Am, 131*(4), 2987-2998.

Chen, J., Baer, T., & Moore, B. C. (2013). Effect of spectral change enhancement for the hearing impaired using parameter values selected with a genetic algorithm. *J Acoust Soc Am, 133*(5), 2910-2920.

Chung, K. (2004). Challenges and recent developments in hearing aids: Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends in Amplification, 8*(3), 83-124.

Cooke, M., Garcia Lecumberri, M., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America, 123*(1), 414-427.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *J Acoust Soc Am, 123*(1).

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America, 24*(6), 597-606.

Corey, R. M., Jones, U., & Singer, A. C. (2020). Acoustic effects of medical, cloth, and transparent face masks on speech signals. *The Journal of the Acoustical Society of America, 148*(4), 2371-2375.

Crandell, C. C., & Smaldino, J. J. (1996). Speech perception in noise by children for whom English is a second language. *American Journal of Audiology, 5*(3), 47-51.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication, 47*(1-2), 32-42.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America, 116*(6), 3668-3678.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics, 22*(2), 109-122.

Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.

Durand, J. (2005). Tense/lax, the vowel system of English and phonological theory. *Headhood, elements, specification and contrastivity: Phonological papers in honor of John Anderson. Philadelphia, PA: John Benjamins*, 77-98.

Durlach, N. I., Mason, C. R., Kidd Jr, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *The Journal of the Acoustical Society of America, 113*(6), 2984-2987.

Egan, J. P., & Hake, H. W. (1950). On the masking pattern of a simple auditory stimulus. *The Journal of the Acoustical Society of America, 22*(5), 622-630.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37*(4), 452-465.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America, 88*(4), 1725-1736.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437-470.

Franck, B. A., van Kreveld-Bos, C. S., Dreschler, W. A., & Verschuure, H. (1999). Evaluation of spectral enhancement in hearing aids, combined with phonemic compression. *J Acoust Soc Am, 106*(3 Pt 1), 1452-1464.

Franck, B. A., van Kreveld-Bos, C. S. G., Dreschler, W. A., & Verschuure, H. (1999). Evaluation of spectral enhancement in hearing aids, combined with phonemic compression. *The Journal of the Acoustical Society of America, 106*(3), 1452-1464.

Gat, I. B., & Keith, R. W. (1978). An effect of linguistic experience: Auditory word discrimination by native and non-native speakers of English. *Audiology, 17*(4), 339-345.

Giannakopoulou, A. (2012). *Plasticity in second language (L2) learning: perception of L2 phonemes by native Greek speakers of English* (Doctoral dissertation, School of Social Sciences Theses).

Goldin, A., Weinstein, B., & Shiman, N. (2020). How do medical masks degrade speech perception. *Hearing review, 27*(5), 8-9.

Guan, J., & Liu, C. (2019a). Speech Perception in Noise With Formant Enhancement for Older Listeners. *J Speech Lang Hear Res, 62*(9), 3290-3301.

Guan, J., & Liu, C. (2019b). Vowel discrimination in noise with formant enhancement: Effects of hearing loss and aging. *The Journal of the Acoustical Society of America, 146*(4), 2922-2922.

Guan, J., Liu, C., Tao, S., Li, M., Wang, W., & Dong, Q. (2015). Sentence recognition in temporal modulated noise for native and non-native listeners: Effect of language experience. *The Journal of the Acoustical Society of America, 137*(4), 2383-2383.

Guan, J., Liu, C., Tao, S., Mi, L., Wang, W., & Dong, Q. (2015). Vowel identification in temporal-modulated noise for native and non-native listeners: Effect of language experience. *The Journal of the Acoustical Society of America, 138*(3), 1670-1677.

Gustafsson, H. Å., & Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *The Journal of the Acoustical Society of America, 95*(1), 518-529.

Halle, M., Hughes, G. W., & Radley, J. P. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America, 29*(1), 107-116.

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Effect of third-formant transitions on the perception of the voiced stop consonants. *The Journal of the Acoustical Society of America, 30*(2), 122-126.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J Acoust Soc Am, 97*(5 Pt 1), 3099-3111.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America, 109*(2), 748-763.

Holder, J. T., Levin, L. M., & Gifford, R. H. (2018). Speech recognition in noise for adults with normal hearing: Age-normative performance for AzBio, BKB-SIN, and QuickSIN. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 39*(10), e972.

Hsieh, B. R., & Pan, H. H. (2010, December). L2 experience and non-native vowel categorization of L1-mandarin speakers. In *11th Annual Conference of the International Speech Communication Association: Spoken Language Processing for All, INTERSPEECH 2010*.

Hu, W., Mi, L., Yang, Z., Tao, S., Li, M., Wang, W., . . . Liu, C. (2016). Shifting Perceptual Weights in L2 Vowel Identification after Training. *PloS one, 11*(9), e0162876.

Jin, S. H., & Liu, C. (2012). English sentence recognition in speech-shaped noise and multi-talker babble for English-, Chinese-, and Korean-native listeners. *J Acoust Soc Am, 132*(5), EL391-397. doi:10.1121/1.4757730

Kahneman, D. (1973). *Attention and effort* (Vol. 1063, pp. 218-226). Englewood Cliffs, NJ: Prentice-Hall.

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *J Acoust Soc Am, 72*(2), 379-389.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In *Auditory perception of sound sources* (pp. 143-189): Springer.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *J Acoust Soc Am, 124*(6), 3959.

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*: Cengage learning.

Lecumberri, M. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America, 119*(4), 2445-2454.

Leung, K. K., Jongman, A., Wang, Y., & Sereno, J. A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America, 140*(1), 45-58.

Li, M., Wang, W., Tao, S., Dong, Q., Guan, J., & Liu, C. (2016). Mandarin Chinese vowel-plus-tone identification in noise: Effects of language experience. *Hearing research, 331*, 109-118.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied, 68*(8), 1.

Lindblom, B. E., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *J Acoust Soc Am, 42*(4), 830-843.

Lipski, S. C., Escudero, P., & Benders, T. (2012). Language experience modulates weighting of acoustic cues for vowel perception: an event-related potential study. *Psychophysiology, 49*(5), 638-650.

Liu, C., & Eddins, D. A. (2008). Effects of spectral modulation filtering on vowel identification. *J Acoust Soc Am, 124*(3), 1704-1715.

Liu, C., & Jin, S.-H. (2019). Psychometric Functions of Vowel Detection and Identification in Long-Term Speech-Shaped Noise. *Journal of Speech, Language, and Hearing Research, 62*(5), 1473-1485.

Liu, C., & Kewley-Port, D. (2004). Formant discrimination in noise for isolated vowels. *The Journal of the Acoustical Society of America, 116*(5), 3119-3129.

Liu, C., Tao, S., Wang, W., & Dong, Q. (2012). Formant discrimination of speech and non-speech sounds for English and Chinese listeners. *The Journal of the Acoustical Society of America, 132*(3), EL189-EL195.

Lock, D. (2009). The New Children's Encyclopedia. In: DK Publishing, Inc, New York, NY.

Luo, C. L. (2002). Production and perception of similar and new vowels by Mandarin speakers: An experimental study of English high vowels. Unpublished doctoral dissertation, National Taiwan Normal University.

Lyzenga, J., Festen, J. M., & Houtgast, T. (2002). A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity. *J Acoust Soc Am, 112*(3 Pt 1), 1145-1157

Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *The Journal of the Acoustical Society of America, 63*(2), 572-580.

Mi, L., Tao, S., Wang, W., Dong, Q., Guan, J., & Liu, C. (2016). English vowel identification and vowel formant discrimination by native Mandarin Chinese-and

native English-speaking listeners: The effect of vowel duration dependence. *Hearing research, 333*, 58-65.

Mi, L., Tao, S., Wang, W., Dong, Q., Jin, S.-H., & Liu, C. (2013). English vowel identification in long-term speech-shaped noise and multi-talker babble for English and Chinese listeners. *The Journal of the Acoustical Society of America, 133*(5), EL391-EL397.

Miller, G. A. (1947). The masking of speech. *Psychological bulletin, 44*(2), 105.

Miller, G. A., & Licklider, J. C. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America, 22*(2), 167-173.

Miller, R. L., Calhoun, B. M., & Young, E. D. (1999). Contrast enhancement improves the representation of/ε/-like vowels in the hearing-impaired auditory nerve. *The Journal of the Acoustical Society of America, 106*(5), 2693-2708.

Morrison, G. S. (2009). L1-Spanish speakers' acquisition of the English /i/-/I/ contrast II: perception of vowel inherent spectral change. *Lang Speech, 52*(Pt 4), 437-462.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and speech, 36*(1), 39-66.

Nábělek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *The Journal of the Acoustical Society of America, 75*(2), 632-634.

Olive, J. P., Greenwood, A., & Coleman, J. (1993). *Acoustics of American English speech: A dynamic approach*: Springer Science & Business Media.

Parikh, G., & Loizou, P. C. (2005). The influence of noise on vowel and consonant cues. *J Acoust Soc Am, 118*(6), 3874-3888.

Peterson, G. E., & Barney, H. L. (1951). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 23*(1), 148-148.

Phatak, S. A., & Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *J Acoust Soc Am, 121*(4), 2312-2326.

Plomp, R. (2019). Perception of speech as a modulated signal. In *Proceedings of the tenth international congress of phonetic sciences* (pp. 29-40). De Gruyter Mouton.

Plomp, R., & Mimpen, A. (1979a). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18*(1), 43-52.

Plomp, R., & Mimpen, A. M. (1979b). Speech-reception threshold for sentences as a function of age and noise level. *J Acoust Soc Am, 66*(5), 1333-1342. doi:10.1121/1.383554

Rosenhouse, J., Haik, L., & Kishon-Rabin, L. (2006). Speech perception in adverse listening conditions in Arabic-Hebrew bilinguals. *International Journal of Bilingualism, 10*(2), 119-135.

Rout, A. (2006). *The effect of spectral enhancement on speech recognition performance of normal-hearing and hearing-impaired individuals.* Purdue University,

Simpson, A. M., Moore, B. C. J., & Glasberg, B. R. (1990). Spectral Enhancement to Improve the Intelligibility of Speech in Noise for Hearing-impaired Listeners. *Acta Otolaryngol, 109*(sup469), 101-107.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America, 118*(5), 2775-2778.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America, 64*(5), 1358-1368.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America, 55*(3), 653-659.

Stone, M. A., & Moore, B. C. (1992). Spectral feature enhancement for people with sensorineural hearing impairment: effects on speech intelligibility and quality. *Journal of rehabilitation research and development, 29*(2), 39-56.

Story, B. H., & Bunton, K. (2010). Relation of Vocal Tract Shape, Formant Transitions, and Stop Consonant Identification. *Journal of Speech, Language, and Hearing Research, 53*, 1514-1528..

Strange, W. (1989). Evolving theories of vowel perception. *The Journal of the Acoustical Society of America, 85*(5), 2081-2087.

Stuart, A., Zhang, J., & Swink, S. (2010). Reception thresholds for sentences in quiet and noise for monolingual English and bilingual Mandarin-English listeners. *J Am Acad Audiol, 21*(4), 239-248.

Summerfield, Q., Foster, J., Tyler, R., & Bailey, P. J. (1985). Influences of formant bandwidth and auditory frequency selectivity on identification of place of articulation in stop consonants. *Speech Communication, 4*(1-3), 213-229.

Takata, Y., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *The Journal of the Acoustical Society of America, 88*(2), 663-666.

Tao, S., Chen, Y., Wang, W., Dong, Q., Jin, S.-H., & Liu, C. (2018). English Consonant Identification in Multi-Talker Babble: Effects of Chinese-Native Listeners' English Experience. *Language and speech*, 002383091879060.

Ter Keurs, M., Festen, J. M., & Plomp, R. (1992). Effect of spectral envelope smearing on speech reception. I. *The Journal of the Acoustical Society of America, 91*(5), 2872-2880.

Ter Keurs, M., Festen, J. M., & Plomp, R. (1993). Effect of spectral envelope smearing on speech reception. II. *The Journal of the Acoustical Society of America, 93*(3), 1547-1552.

Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., . . . Kent, R. D. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America, 137*(5), 3005-3007.

Titze, I. R., & Martin, D. W. (1998). Principles of Voice Production. *Acoustical Society of America Journal, 104*(3), 1148.

Turner, C. W., Fabry, D. A., Barrett, S., & Horwitz, A. R. (1992). Detection and recognition of stop consonants by normal. Hearing and hearing impaired listeners. *Journal of Speech, Language, and Hearing Research, 35*(4), 942-949.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America, 126*(1), 367-376.

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first-and second-language multi-talker babble. *Speech Communication, 52*(11-12), 943-953.

Vento, B., & Durrant, J. (2009). Handbook of clinical audiology. *Assessing bone conduction thresholds in clinical practice, 6th edn. Williams & Wilkins, Baltimore*.

Wang, H. (2007). English as a lingua franca: mutual intelligibility of Chinese, Dutch and American speakers of English (*Doctoral dissertation, Leiden University).*

Wang, M. (2017). Variation of Vowels when Preceding Voiced And Voiceless Consonant in Sundanese. *International Refereed Journal of Engineering and Science, 6*(9), 13-20.

Wang, X. (2006). Mandarin listeners' perception of english vowels: Problems and strategies. *Canadian Acoustics, 34*(4), 15-26.

Wegel, R., & Lane, C. (1924). The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Physical review, 23*(2), 266.

Woodall, A., & Liu, C. (2013). Effects of Signal Level and Spectral Contrast on Vowel Formant Discrimination for Normal-Hearing and Hearing-Impaired Listeners. *American Journal of Audiology, 22*(1), 94-104.

Yang, J., Luo, F.-L., & Nehorai, A. (2003). Spectral contrast enhancement: Algorithms and comparisons. *Speech Communication, 39*(1-2), 33-46.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of cognitive neuroscience, 22*(6), 1319-1332.